Light Field Image Compression and Compressive Acquisition

Fatma Hawary

THESE DE DOCTORAT DE

L'UNIVERSITE DE RENNES 1
COMUE UNIVERSITE BRETAGNE LOIRE

ECOLE DOCTORALE N° 601
Mathématiques et Sciences et Technologies de l'Information et de la Communication
Spécialité : Traitement du signal

Par

Fatma HAWARY

Light Field Image Compression and Compressive Acquisition

Thèse présentée et soutenue à Rennes, le 29 Mai 2019
Unité de recherche : INRIA Rennes – Bretagne Atlantique et Technicolor R&I

Rapporteurs avant soutenance :

André Kaup — Professor, Friedrich-Alexander University - Erlangen-Nürnberg, Germany
Mårten Sjöström — Professor, Mid Sweden University, Sweden

Composition du Jury :

Président :	Atanas Gotchev	Professor, Tampere University of Technology, Finland
Examinateur :	Laurent Albera	University Lecturer, ESIR/ISTIC, Université de Rennes 1, France
Dir. de thèse :	Christine Guillemot	Research Director, INRIA, France
Co-dir. de thèse :	Guillaume Boisson	Senior Scientist, Technicolor, France

Acknowledgements

Throughout the work on this thesis, I have received a great deal of support and assistance.

First and foremost, I would like to thank my supervisors, Dominique Thoreau, Guillaume Boisson and Christine Guillemot who offered their advice and helped with their invaluable expertise throughout the journey. Their methodical guidance and great efforts were essential to the accomplishment of this thesis.

I would like to thank my dissertation committee members, namely Prof. Marten Sjöström, Prof. André Kaup, Prof. Atanas Gotchev and Dr. Laurent Albera for their crucial comments and suggestions that shaped my final thesis.

I would also like to thank my colleagues from Technicolor, for the much pleasant work environment they offered and their valuable support during the last three years.
My time at Technicolor was made much enjoyable thanks to the kind people I was happy to meet. Although the list is far from being exhaustive, I would like to acknowledge Valérie, Matthieu, Rémy, Neus, Tristan, Benoit, Paul, Jean, Dmitry, Franck H., Christophe, Thierry, Mozhdeh and Erik. I am thankful for all the incessant support they brought during my thesis. Thank you.

Special thanks go to the Sirocco team at Inria for their warm welcome, their valuable help and lively discussions. I further acknowledge my labmates: Mira, Lara, Navid, Mai, Maja, Elian, Pierre D., Pierre A., Thierry, Simon, Xiaoran and Jinglei for their support and collaboration. Thank you.

A heartfelt thank you goes to my friends: Salma, Hadrien, Anas, Dhouha, Amal, Hamdi and Amina. Your incessant support and much love have made this journey much more special. Thank you so much.

Last but not least, I would like to acknowledge the people who mean the world to me, my parents. There are no words to describe how grateful I am. Thank you for your selfless love, care and dedicated efforts that contributed to this achievement.

To my parents.

Contents

Résumé en Français

Contexte

Dans plusieurs domaines utilisant des contenus de type image ou vidéo, il y a un intérêt particulier à fournir une représentation fidèle des scènes capturées, que ce soit en donnant une impression de profondeur ou de géométrie tridimensionnelle dans les contenus diffusés. Durant les dernières décennies, de nombreux domaines de recherche et de l'industrie se sont penchés sur le problème de capture des données offrant le rendu le plus naturel possible à l'utilisateur. Ceci est aussi dans le but d'assurer des expériences immersives plus réussies.

Puisque les images 2D donnent une représentation plate des scènes du monde réel qui est lui tridimensionnel, la vision 3D est devenue un domaine de recherche très actif au cours des dernières annéees. En effet, les évolutions récentes dans l'industrie de l'image offrent de plus en plus de données sur la profondeur et la géométrie de la scène, ainsi que des images de haute résolution permettant d'avoir encore plus de détails pour reconstruire la scène 3D. De plus, la nouvelle ère de la réalité augmentée ouvre de nouveaux champs d'intérêt pour la reconstruction 3D, et requiert des acquisitions d'images à large échelle.

Alors que les technologies Ultra High Definition (UHD), avec les résolutions 4K et 8K, les technologies High Dynamic Range (HDR) et White Color Gamut (WCG) permettent aux vidéos 2D d'atteindre les limites de perception les plus élevées du système visuel humain, les technologies vidéo 3D actuelles ont des difficultés à conquérir le marché consommateur en raison de limitations techniques, mais aussi du fait qu'elles n'apportent pas encore un niveau de confort visuel acceptable.

Par exemple, la 3D stéréoscopique n'utilise que deux vues et ne permet pas à l'utilisateur de changer de point de vue. La perception de la profondeur, élément clé des applications immersives, n'est pas non plus garantie dans ce cas.

Les images de type *light field* (ou champs de lumière) ont récemment gagné en popularité, à la fois dans les domaines académiques et industriels. Des efforts ont été entrepris pour explorer le potentiel de nouveaux dispositifs et formats de champs de lumière, tels que JPEG Pleno [1] ou "the Joint ad hoc group for digital representations of light/sound fields for immersive media applications" [2]. En effet, les images de type *light field* permettent une variété d'applications telles que l'estimation de profondeur, le refocusing, la super-résolution et la reconstruction 3D. Cependant, l'immense volume de données capturées par les champs de lumière représente encore

[1] https://jpeg.org/jpegpleno

[2] https://mpeg.chiariglione.org/standards/mpeg-i/technical-report-immersive-media/report-joint-ad-hoc-group-digital-0

un défi à la fois en termes d'acquisition, de stockage et de transmission.

Même en présence de techniques permettant de synthétiser davantage de vues à partir d'un nombre réduit de vues, des artefacts peuvent survenir, résultant soit du dispositif d'acquisition, ou des problèmes de codage de la profondeur. En effet, plusieurs aspects de la scène capturée, tels que les occlusions, les spécularités et les variations de profondeur, doivent être traités afin de générer un champ de lumière représentant une large parallaxe.

Alors que de nombreuses méthodes compressent efficacement les champs de lumière, généralement capturée avec une caméra *plénoptique* en exploitant les corrélations entre images à l'aide de techniques basées sur la compensation de disparité, les performances de compression d'images de champ de lumière à grande échelle avec des parallaxes importantes sont encore limitées.

Les champs de lumière présentent d'autre part des structures qui peuvent être exploitées dans les algorithmes de compression en utilisant des modèles adaptés pour représenter ce type de données.

Motivations et Objectifs

Un champ de lumière peut être vu comme une représentation des intensités d'un ensemble de rayons lumineux dans la scène capturée. D'un point de vue général, le champ de lumière peut faire référence aux captures multiples d'une scène à partir de plusieurs points de vue. En ce sens, de nombreux contenus peuvent correspondre à cette définition. Les images d'une vidéo d'une caméra en mouvement, ou même deux images d'une caméra de type *stéréo* peuvent techniquement former un champ de lumière. Cependant, nous supposons ici que la structure du champ de lumière est telle que tous les points de vue sont placés dans le même plan, avec une distance régulière entre chaque paire de points de vue adjacents. Nous considérons le cas du champ de lumière contenant plus que deux vues. Avec ces hypothèses, un champ de lumière contient un grand volume d'informations sur la scène qui peut être utilisé dans de nombreuses applications de traitement de contenus images.

Cependant, la large quantité de données dans un champ lumière présente un souci critique, vu les limitations en termes de capacités de transmission actuelles. Bien que les codecs existants offrent des performances assez intéressantes pour la compression d'images et de vidéos 2D, et même en présence des extensions multi-vues des standards de codage, il est néanmoins nécessaire d'améliorer les schémas de compression pour mieux les adapter aux données de type champs de lumière.

En effet, même si un champ de lumière contient des images prises à différents points de vue, leur disposition régulière engendre des redondances structurées. Cela motive l'utilisation de ces redondances pour réduire les coûts de codage. Une solution possible consiste à choisir seulement quelques échantillons du champ de lumière pour prédire le champ de lumière tout entier.

Un autre aspect intéressant à traiter dans les champs de lumière consiste à améliorer la résolution finale du champ de lumière à partir d'une acquisition de résolution beaucoup plus réduites. En effet, avec l'intérêt croissant porté aux applications de vision 3D telles que la réalité virtuelle et les images 360°, il est de plus en plus nécessaire de fournir des champs de lumière avec une large résolution spatiale mais aussi angulaire. Mais vu les limites actuelles en termes de puissance de traitement et outils de stockage, la génération de données de plus en plus volumineuses en champ de lumière ne peut pas être directe. Dans ce contexte, une

acquisition compressée s'avère être une solution intéressante pour restaurer avec précision la haute dimensionnalité des données de type *light field*. Il s'agit dans ce cas de ne capturer qu'un nombre réduit de mesure à partir de laquelle le champ de lumière correspondant pourrait être complètement reconstruit.

L'objectif principal de cette thèse est d'explorer la compressibilité des contenus de type champs de lumière dans divers contextes. L'étude est réalisée dans le cadre de la compression d'image, l'efficacité de la compression étant un facteur clé pour permettre l'utilisation de ces contenus à la fois sur le marché du grand public et dans l'industrie.

Nous étudions d'abord les différentes technologies et méthodes d'acquisition et de compression des champs de lumière. Nous proposons ensuite de nouvelles solutions de compression de ces données, d'abord pour permettre une transmission efficace de ces contenus à grand volume, ensuite pour faciliter l'acquisition de larges champs de lumière, et ainsi répondre aux besoins de capture d'images à plus grande échelle surtout pour la vision 3D.

Résumé des Contributions

Cette thèse est structurée en deux parties principales.
Dans la partie I, le **chapitre 1** fournit une introduction globale au domaine de l'imagerie de type *light field* (ou champs de lumière). Plus précisément, nous introduisons les notions de base en termes de formulation de l'image *light field*, sa structure et ses propriétés. Nous présentons également les différentes applications que les champs de lumière permettent comme l'estimation de profondeur ou le *refocusing* (refocalisation).
Le **chapitre 2** présente les approches de l'état de l'art pour la compression des images *light field*, ainsi que leur acquisition "compressive". On introduit également dans ce chapitre la théorie derrière l'imagerie "compressive". Cela donne un aperçu général des différentes notions exploitées tout au long de la thèse dans l'étude de compression des données *light field*.

La partie II présente nos différentes contributions. En résumé, cette partie est constituée selon deux axes principaux correspondant d'abord à la compression des images *light field* pour la transmission, et ensuite à l'acquisition et reconstruction de données *light field* sous-échantillonnées. Plus en détail, les contributions sont organisées comme suit:

Chapitre 3: Dans le domaine des images *light field*, il y a peu de travaux sur les nouvelles représentations de données et les techniques adaptées de compression. Par conséquent, les possibilités pour développer des normes consacrées à ce contenu spécifique demandent encore plusieurs recherches. Dans cette première contribution, nous nous intéressons à la représentation du *light field* en images de type "sous-ouverture" qui donnent chacune une vue 2D d'un point de vue de la scène, afin d'exploiter la redondance entre elles. Nous avons développé un premier schéma de compression (compatible avec les codeurs classiques, par exemple HEVC). Le but consiste à minimiser le coût de codage d'une matrice de vues (images de "sous-ouverture"), en ne prenant qu'un nombre réduit des vues, tout en assurant une bonne qualité de compression. L'idée repose sur un schéma de compression scalable basé SHVC. Dans un premier temps, un nombre réduit d'échantillons (vues de la matrice *Light Field*) est encodé en une couche de base; une fois ces images décodées, les vues manquantes seront ensuite synthétisées dans le domaine de Fourier 4D, pour reconstituer la matrice de vues de la couche de base. Cette matrice est ensuite utilisée

comme référence, en termes de prédiction inter-couches, pour encoder le *light field* via une couche d'amélioration. Nous avons étudié les différentes étapes de ce schéma de compression et évalué sa performance. La méthode de référence utilisée pour la comparaison réside dans le codage en une seule couche (*single layer*), de toutes les vues de la matrice placées dans une pseudo-séquence à l'entrée du codeur HEVC. L'approche proposée offre de bonnes performances moyennant une réduction de débit de 11.2% (pouvant atteindre les 25% pour des champs de lumière à large parallax) par rapport à la méthode de référence. Par ailleurs, l'impact de la compression a aussi été évalué sur une application *light field*, la synthèse d'image *all-in-focus*. Ces évaluations ont montré l'avantage de notre méthode comparée à la méthode de référence HEVC *single layer*.

La méthode proposée est efficace, peut être utilisée pour différents contenus de champs de lumière, *sparses* ou denses, et même avec une grande parallaxe.

Chapitre 4: Les travaux présentés dans ce chapitre sont liés à la problématique d'acquisition des données *light field*, notamment aux limites en termes de résolution spatio-angulaire des dispositifs d'acquisition actuels. En effet, ces appareils fournissent généralement un champ de lumière basse résolution en multiplexant plusieurs vues sur une seule image de capteur 2D, ou alors nécessitent l'acquisition de plusieurs images successives pour générer un champ de lumière haute résolution. Des travaux dans le domaine du *Compressive Sensing* ont cependant montré qu'un signal peut être restauré à partir d'un nombre réduit d'échantillons, à la condition que ce signal soit parcimonieux dans un certain espace de représentation. Les images réelles de type *light field* pouvant être parcimonieusement représentés dans le domaine de Fourier, un signal *light field* sous-échantillonné pourrait donc être correctement reconstruit.

Nous avons développé une méthode de reconstruction d'un champ de lumière à partir d'un signal original échantillonné aléatoirement, opérant dans le domaine de transformée de Fourier 4D. Pour reconstruire le signal 4D, l'algorithme sélectionne itérativement les fréquences qui permettent de mieux modéliser les échantillons connus (en termes de distorsion). L'idée est inspirée d'une méthode nommée FSR (Frequency Selective Reconstruction) [1] et étendue au cas des signaux *light field* 4D. La reconstruction est réalisée par hyper-bloc 4D, chaque hyper-bloc étant considéré comme le noyau d'un voisinage spatio-angulaire. Le modèle d'approximation est construit en ajoutant à chaque itération une nouvelle fonction de base, par minimisation de l'erreur résiduelle. L'approche est semblable à la méthode du Matching Pursuit [2] dans le sens où elle vise à approximer un signal en choisissant les meilleures projections de ce signal dans la base de fonctions de Fourier. Néanmoins, l'approche FSR ne garantit pas l'orthogonalité du résidu par rapport à la base de toutes les fréquences déjà sélectionnées. En conséquence, une fréquence peut être choisie plus fois, et donc, un nombre plus important d'itérations est nécessaire pour estimer correctement la contribution de chaque fréquence. Une première amélioration a été apportée à cette méthode afin d'assurer l'orthogonalité du résidu, en prenant en compte l'impact de l'ajout de chaque fréquence sur les coefficients des fonctions de base précédemment sélectionnées. Un raffinement des fréquences sélectionnées a également été proposé afin de s'approcher plus du spectre de Fourier continu du *light field* 4D. L'algorithme proposé a été implémenté et validé sur différents types de données *light field*. La qualité des images reconstruites dépasse les 34dB même pour un échantillonnage assez réduit de l'ordre de 4%. Comparée à plusieurs méthodes de l'état de l'art, notre approche permet également d'obtenir des gains significatifs en qualité de reconstruction en termes de PSNR et SSIM.

Chapitre 5: Dans les précédents chapitres, nous avons proposé des solutions pour l'acquisition et la transmission des données *light field*. Ces approches exploitent essentiellement la sparsité du champ de lumière dans le domaine de transformée de Fourier 4D. Dans ce chapitre, nous proposons une étude expérimentale sur l'impact du schéma complet comprenant les deux approches sur la qualité finale d'un champ de lumière. L'étude est menée en variant pour chaque étape le taux de compression. Le but étant de proposer un système de traitement des images *light field* qui englobe l'acquisition ainsi que la transmission du contenu *light field*. Nous menons plusieurs expériences afin de mesurer la variation de qualité et de conclure sur les paramètres garantissant un niveau de qualité d'image minimum correct. Nous comparons également notre schéma aux méthodes basées sur les disparités afin de déduire les différences et les avantages de ces approches différentes.

Pour conclure, les contributions de cette thèse se sont principalement concentrées sur les techniques de compression des données *light field* pour apporter un contenu de plus grande résolution et assurer de meilleures performances en terme de codage.

Plusieurs améliorations peuvent être envisagées pour les approches présentées dans cette thèse. Le schéma de codage scalable pourrait être plus adaptable en rendant la méthode de reconstruction des vues manquantes dans la couche de base plus flexible au niveau du choix des échantillons angulaires. En outre, avec l'arrivée imminente des premiers modèles du nouveau standard de codage H266/VVC, une comparaison de performance et de complexité avec le schéma proposé pourrait être envisagée. Les solutions proposées ici pourraient également être étendues aux vidéos light field, en ajoutant la dimension temporelle dans le modèle du champ de lumière, et en utilisant les corrélations entre les images pour assurer de bonnes performances en terme de compression.

En ce qui concerne l'acquisition "compressive" des données de champ de lumière, plusieurs améliorations pourraient être envisagées. Tout d'abord, une étude approfondie peut être réalisée pour déterminer le modèle d'échantillonnage optimal à appliquer à chaque vue.

Déterminer le taux d'échantillonnage théorique minimal pour assurer une reconstruction correcte du light field serait également envisageable, en étudiant la structure compressible des light fields, ce qui permet d'adapter le seuil théorique proposé dans le contexte du "compressive sensing" [3] aux données de type *light field*.

Introduction

Context

Providing a faithful representation of real scenes through images and videos has always been a critical need in the imaging research and industry. Seeking image fidelity has been one of the principal objectives in many application areas. Since 2D images are a flat representation of the 3D world scenes, 3D and depth imaging has become a highly active research area in the last decades. Indeed, recent evolutions in the image industry offer more and more depth range data and high resolution photographs, and the upcoming era of augmented reality is opening new fields of interests in 3D reconstruction.

Lately, the recent progress in video technologies tends to provide increasingly immersive experiences. While the Ultra High Definition (UHD), with 4K and 8K resolutions, High Dynamic Range (HDR) and White Color Gamut (WCG) technologies are bringing 2D videos towards the highest perception limits of the human visual system, the current 3D video technologies still struggle to gain the consumer market due to technical limitations but also to insufficient comfortable experiences to the viewer. For instance, stereoscopic 3D only uses two views, and does not allow the viewer to change the point of view, and the perception of depth, which is a key element of immersive applications, is not ensured.

Thus, the next generation of immersive technologies poses major technical challenges, especially with the light field techniques which appear to be very promising solutions for capturing scenes from different perspectives. Indeed, as 2D images and videos provide one view of the scene from a single angle, light fields provide a wider and denser sampling relying on a large number of captured images.

Light field imaging has gradually acquired a significant interest over the last two decades, both in research and industry. Efforts have been undertaken to explore the potential for new light field devices and formats, such as in JPEG Pleno [3] or "the Joint ad hoc group for digital representations of light/sound fields for immersive media applications" [4]. However, the huge amount of data in light fields can represent a critical challenge in terms of acquisition, post-capture applications, as well as future transmission on networks and storage facilities to deploy.

Even in the presence of techniques to synthesize more views from a small number of views and depth, disturbing artefacts may occur, either resulting from the acquisition system or the coding of depth. Indeed, several aspects of the captured scene such as occlusions, specularities and

[3] https://jpeg.org/jpegpleno

[4] https://mpeg.chiariglione.org/standards/mpeg-i/technical-report-immersive-media/report-joint-ad-hoc-group-digital-0

depth variation still need to be dealt with in order to generate a light field with a wide enough baseline. While numerous methods succeed to efficiently compress light fields by exploiting inter-view correlations in disparity compensation-based techniques, the compression performance for high-resolution light field images with large baselines are still limited.

Besides, light field images exhibit structures that can be exploited in the compression algorithms by effectively representing the data in efficient compressive models.

Motivations and goals

A light field represents the intensities of a collection of light rays in the captured scene. From a general point of view, light field can refer to the multiple captures of a scene from several viewpoints. In this sense, many contents can match with this definition. Video frames from a moving camera, or even two images of a stereo camera can technically form a light field. However, we make the assumption here that the light field structure ensures that all viewpoints are placed in the same plane, with a regular distance between each pair of adjacent viewpoints. We also suggest that the light field contains more than two views. With these hypotheses, a light field captures a large volume of information about the scene that can be leveraged for further processing applications.

One challenging issue with the massive amount of data in a light field is related to the transmission capacities that may prove limited in power and speed for such data. While existing codecs provide a significantly high performance for 2D image and video compression, there is still need for improvement of the light field compression schemes. Indeed, even with proposing the multi-view extensions of standard codecs, there is still way to enhance the compression performance and to use new coding solutions more adapted to the light field content. Even if a light field contains varying viewpoints, their regular disposition yields structured redundancies. This motivates the use of these redundancies to reduce the coding cost. A possible way is to choose certain light field samples to predict the remaining ones.

A second interesting aspect is light field acquisition capacities. With the growing interest brought to 3D vision applications such as Virtual Reality, 360 °images, *etc.*..., the need to provide larger numbers of captured viewpoints of the scene has become a critical task, especially with the current limitations processing power and storage tools towards such increasingly large-scale data. In this context, the compressed sensing has proven to be a promising tool to accurately restore the high dimensionality of image data. This motivates exploring the possibility to acquire only a compressed version of the data and subsequently reconstruct the full resolution light field.

The main goal of this thesis is to explore the compressibility of light field content in various contexts. The study is done through the scope of image compression, as compression efficiency is a key factor for enabling the use of these contents on the consumer and industry markets.

We first study which technologies and formats are promising for light field acquisition and compression considering several types of captured data. We then propose a novel coding solution to enable efficient transmission of light field content. We further introduce a compressive reconstruction solution to facilitate the acquisition of high-resolution light fields, in order to meet the increasing needs for larger scale contents in 3D-vision related domains.

Thesis structure

The dissertation is structured into two main parts.

Part I introduces a general understanding of the context and issues addressed during this thesis.

Chapter 1 provides an introduction to light field imaging. The basic notions related to light field image acquisition and processing are presented, as well as key characteristics of light fields and a broad record of interesting applications related to this 4D content.

Chapter 2 presents in details the state of the art in light field compression and compressed acquisition.

Part II presents our contributions on novel solutions for light field data compression and compressed sensing and reconstruction.

Chapter 3: In this first contribution, we present a scalable compression scheme for light field images. The input data is a subset of light field sub-aperture images, encoded using a standard codec and used to reconstruct the full light field in the Fourier domain. The proposed method is efficient, does not require any prior knowledge of the scene geometry and can be utilized for different light field contents, sparse or dense but also with large baselines. The approach is tested on several light fields, and the decoded images exhibit a good quality. Significant bit-rate reductions of 11.2% in average (but reaching up to 25% for large multi-view light fields) are obtained compared to the baseline method, with any range of disparity and spatial/angular resolutions.

Chapter 4: This chapter tackles the problem of 4D light field compressive acquisition. The objective is to be able to reconstruct a full 4D light field from a sampled version of the data. The sampling rate is pre-fixed, and thus can adapt to the use case or the application demands. A sparse frequency selection in the Fourier domain permits to reconstruct a randomly sampled light field. The sparse 4D spectrum of the light field is restored from the input samples, and a simple inverse Fourier transform provides the full light field. The proposed compressive scheme achieves high quality reconstruction of the sampled light fields, with PSNR values over 34dB even in low compression rates of order of 4%. It offers the possibility to acquire high dimensional light field content, and thus overcome limitations in the existing storage facilities. The algorithm has been improved and optimized for a more accurate approximation in the continuous Fourier domain.

Chapter 5: The previous chapters are related to two main aspects of light field signal processing, and proposed solutions for efficient compression and reconstruction of light fields in both the acquisition and transmission sides. In order to evaluate the impact of the several processing on the quality of the final light field, we propose in this chapter to study the performance of the

full end-to-end scheme from the capture to the broadcast of light field data. We conduct several experiments in order to measure the quality variation and conclude on the parameters that ensure a minimum correct image quality level. We also compare our scheme to disparity-based methods to deduce the differences and advantages of all these different approaches.

Part I

Context and Background

Chapter 1

Light Field Imaging

1.1 Formulation

A fundamental paper introducing the concept of light fields is "*The Plenoptic Function and Elements of Early Vision*" of Adelson and Bergen [4]. They develop the theory of the plenoptic function from the observation that the available information of a scene is contained in the light filling the space. The light is composed of rays that carry information in a form of intensity or radiance. Although light fields can be defined in different systems and types, they fundamentally represent the 3D scene as a collection of light rays through all points in the space and in all directions. In this context, the plenoptic function that usually describes a light field is defined as

$$P(X, Y, Z, \theta, \alpha, \lambda, t). \tag{1.1}$$

It describes a radiance value with rays parametrization of essentially 5 components that can be extended to 7:

- 3D Space coordinates (X,Y,Z),
- a direction defined by polar coordinates (θ, α),
- the time t and the light wavelength λ.

Considering a static scene, the plenoptic function can be reduced to 6 coordinates, excluding t. Also, to represent the color in RGB system, λ can be decomposed to 3 components: Red-Green-Blue.

In sum, the plenoptic function can serve as a general framework to all possible scene imaging modalities, and can be adapted to the imaging device. Furthermore, the plenoptic function can be simplified into four coordinates following a two-plane representation [5] shown in Figure 1.1.

Capturing all the light filling the space within a scene yields to highly redundant information, and is practically infeasible. Thus, all imaging techniques only capture sparse samples of this general function. Several representations of light fields have been proposed in the literature [5–7]. One of the common ways to define a light field is to consider it as collection of pinhole views from several points parallel to a common image plane (see Figure 1.1), known as the *Lumigraph* representation [5].

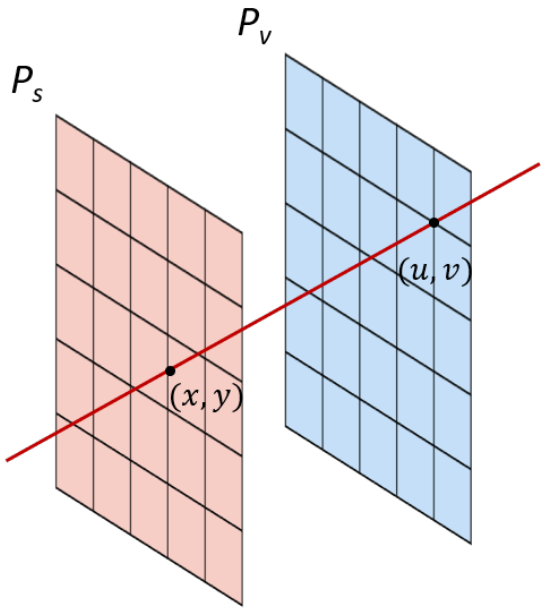


Figure 1.1: 2D-plan representation of the light field function.

It offers a simple way to represent a ray by its intersection with two distinct and parallel planes. The 2D plane P_v contains the focal points of the views that will be parameterized by (u, v) and the image plane P_s by coordinates (x, y). The 4D light field then maps the ray passing through the 4 coordinates (x, y, u, v) to the corresponding intensity value:

$$L : P_s \times P_v \to \mathbb{R}, \qquad (x, y, u, v) \mapsto L(x, y, u, v) \tag{1.2}$$

By convention, we will designate (u, v) as the angular coordinates, and (x, y) as the spatial coordinates, and we will use this representation from now on when referring to a light field signal.

1.2 Light Field Visualization

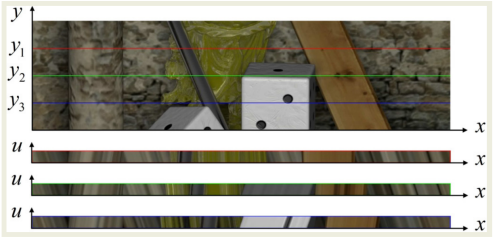
Although the light field function $L(x, y, u, v)$ is a simplified model, it is still hard to imagine a corresponding 4D representation. We present here different ways of visualizing light fields. We can consider the uv plane as a set of cameras with a focal plane at xy. A first configuration in which this model can be seen is that each camera receives the light rays coming from the plane xy and arriving at a point on the uv. Thus, the 4D light field can be represented as a 2D matrix of images, conventionally called *sub-aperture* images (see Figure 1.2). Each sub-aperture image $I_{u*,v*}(x, y)$ is acquired by gathering the samples at the fixed angular coordinates $u*$ and $v*$.

Furthermore, a point on the xy plane represents the light rays bound for all points on the uv plane (the same point seen from different viewpoints). The corresponding image (that can be referred to as *lenslet* image) is formed by gathering, at each point, the rays from different viewpoints.

By gathering the light field samples with one fixed spatial coordinate x and one angular coordinate u (or y and v), one can produce a slice $EP_{x*,u*}(y,v)$ (or $EP_{y*,v*}(x,u)$). The result is called epipolar plane image (EPI) [6]. Unlike a sub-aperture image or a light field lenslet image, the EPI contains information in both spatial and angular dimensions. These images are of particular interest since they make it possible to study the intensity variations, and thus infer the depth and geometry of the scene captured by the light field.

(a) Matrix of sub-aperture images

(b) Lenslet image



(c) Epipolar plane images

Figure 1.2: Visualization of a light field *Buddha* with different ways: (a) sub-aperture images are acquired by gathering the light field samples at fixed angular (u,v) positions, (b) a lenslet image can be acquired by gathering the samples with fixed (x,y) coordinates, (c) epipolar plane images are obtained by fixing the coordinates in both a spatial and an angular dimension.

1.3 Acquisition Systems

A conventional camera averages the intensities of rays described by the higher dimensional plenoptic function into the two-dimensional image sensor of the camera. Light field acquisition devices, however, sample both the angular and spatial dimensions, and differ in the way that this sampling is performed. We can classify these devices into two broad categories.

1.3.1 Camera Arrays

A straightforward way to capture light fields is to move a camera along a 2D direction to sample the 4D ray space. An example is the Stanford camera gantry [1]: a digital camera is mounted on a computer-controlled gantry moving in translation describing the 2D angular plane. This method is only suitable for acquiring static scenes.

An alternative way is to use multiple cameras to capture the same scene at different viewpoints. In this context, camera arrays are composed of synchronized cameras placed along a plane, and often at a regular interval, leading to equally spaced angular samples. Assuming that cameras are identical and similarly placed on a 2D plane, the collection of captured images compose a light field, that can be represented as a matrix of views or sub-aperture images (see an example in Figure 1.2). The distance between each pair of cameras constructs the baseline (*i.e.* angular sampling) and the image resolution provides the spatial sampling.

The camera array views exhibit a certain parallax, due to the large baseline, occlusions may occur and some areas may only be present in a certain number of views. In practice, a camera calibration step is very often needed to rectify the images or to ensure the correspondence between the different views, since misalignment and optical irregularities may appear. The work presented in this thesis assumes the images to be rectified, and does not deal with the camera calibration parameters.

Constructing a camera array is costly and extremely time and effort consuming: it presents technical challenges such as camera calibration/synchronization and requires substantial engineering expertise, which makes these systems rare. Few camera array designs have been introduced to capture light field data. Wilburn *et al.* [8] built a 2D grid composed of 100 video cameras which stream live video to a stripped disk array. Due to the large amount of data generated by this array, the light field is rendered in post processing rather than in real-time. The MIT light field camera array [9] uses a smaller array of 64 commodity video cameras instead of firewire cameras and is capable of synthesizing real-time dynamic light fields. Both systems, however, still suffer from spatial aliasing because of the baseline between neighboring cameras.

More recent camera array-based systems have been proposed in [10,11] using modern cameras capturing high-resolution images (Figure 1.3). The available commercialized systems as presented as service (*e.g.* the *Radius Camera Array System* [2]), or targeted to 360/virtual Reality videos [3]. However, some recent fully synchronizable cameras (such as the *Sony RX0* [4]) are now made accessible for multi-camera shooting.

Finally, multi-camera smartphones providing few (2 to 5) views and other existing small devices can be assimilated to camera arrays.

[1] http://lightfield.stanford.edu/acq.html
[2] www.radiantimages.com
[3] https://www.blog.google/products/google-vr/experimenting-light-fields/
[4] https://www.sony.com/electronics/RX0-series

Figure 1.3: Example of camera array for light field capture: Left: Stanford multi-camera Array [12]. Right: Technicolor's camera array [10] prototype, composed of 16 video cameras.

1.3.2 Plenoptic Cameras

The basic approach on which plenoptic cameras are built is the work of Adelson *et al.* [13] and Ng *et al.* [14]. A plenoptic camera is mainly constructed as classic digital camera with a digital sensor, a main optics and an aperture, at the difference that it encloses also a micro-lens array placed behind the main lens. A plenoptic camera captures a *Lumigraph*. In contrast to a classic camera which integrates the focused light of the main lens on a single sensor element, the micro-lenses in a plenoptic camera split the incoming light by direction of the rays mapping them onto the sensor (see Figure 1.4). The two *Lumigraph* planes correspond hence to the main lens plane and the micro-lens plane.

The resulting images suffer from a spatio-angular trade-off, since the rays arriving into micro-lenses are multiplexed by direction to a single sensor: the total resolution of the image sensor is shared by both spatial and angular resolution, resulting in fairly low resolution micro-lens images. The angular sampling can be considered as dense since the distance between micro-lenses is very small, while the spatial sampling is sparse.

The light field data captured using plenoptic camera consists of a collection of micro-lens images (also named as micro-images). For a matter of visualization, the captured light field is usually rearranged into sub-aperture images. A sub-aperture image is constructed by associating all the co-located pixels from each micro-lens image (*i.e.* at the same angular position) and placing them in a spatial coordinate that corresponds to the micro-lens image they came from.

The matrix of sub-aperture images is similar to the result we get from a camera array, but with much smaller baseline. Their number depends on the number of pixels behind each micro-lens, and their individual resolution corresponds to the number of micro-lenses. These images are of more interest to use for further processing or study of the light field since the structure of

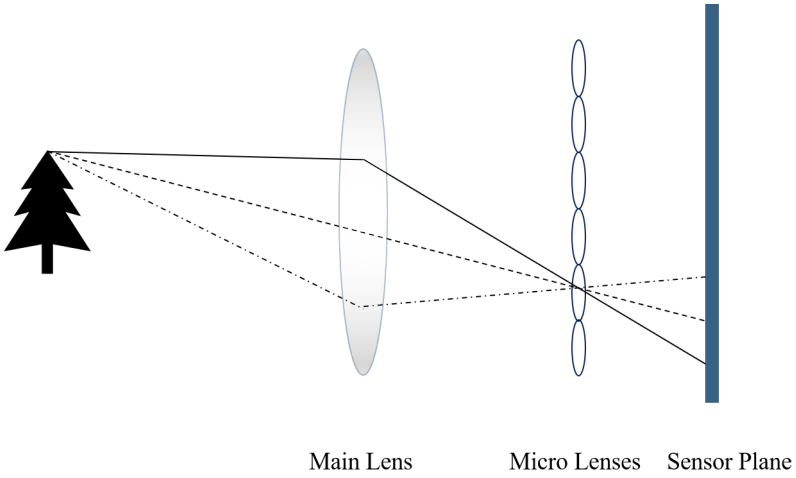


Figure 1.4: A simplified illustration of how a plenoptic camera captures a light field. The angular plane corresponds to the main lens plane and the spatial plane to the micro-lens plane. Each pixel in the micro-lens image corresponds to the same point in the scene.

(a) Lytro 1 (b) Lytro Illum

Figure 1.5: Examples of plenoptic cameras: (a) Lytro 1 camera (b) Lytro Illum camera. They use a micro-lens array to interlace the images of different views on a single sensor.

micro-lens images is hard to exploit.

There are few plenoptic camera designs that have been commercialized. The most known ones are the *Lytro 1* and *Lytro Illum* (Figure 1.5) which are available for consumers, while the Raytrix [5] camera is proposed with high-end design destined for industry.

[5] http://www.raytrix.de

1.4 Light Field Data Properties

By capturing scenes in various viewpoints, light fields exhibit information redundancies and very often views (or sub-aperture images) have high correlations. These redundancies have been often used within various sparsity assumptions for light fields. Indeed, sparsity is one of the main characteristics that has been very often used to propose reconstruction methods of a light field only with a subset of samples (or coded samples).

Many of the interesting properties of light fields come from the frequency domain. Indeed, Fourier analysis of light fields has been extensively used in light transport [15], wave optics [16], as well as for predicting upper bounds on sharpness [17, 18]. In the *Fourier Slice Photography* theorem [17], Ng derived a new algorithm for light field rendering in the frequency domain, deduced from the geometrical optics of image formation: it is based on the Fourier Slice theory that states that in the frequency domain, an image formed with a full lens aperture is a 2D slice in the 4D light field. In other words, the images focused at different depths correspond to slices at different angles/trajectories in the 4D space. This mathematical theory was exploited to analyze the performance of digital refocusing, and further leaded to a Fourier-based approach for digital refocusing of light fields. Furthermore, Zhang and Levoy [16] explored the correspondence between the Fourier slice photography in computer graphics and wave-front coding in optics using the Wigner distribution and light fields.

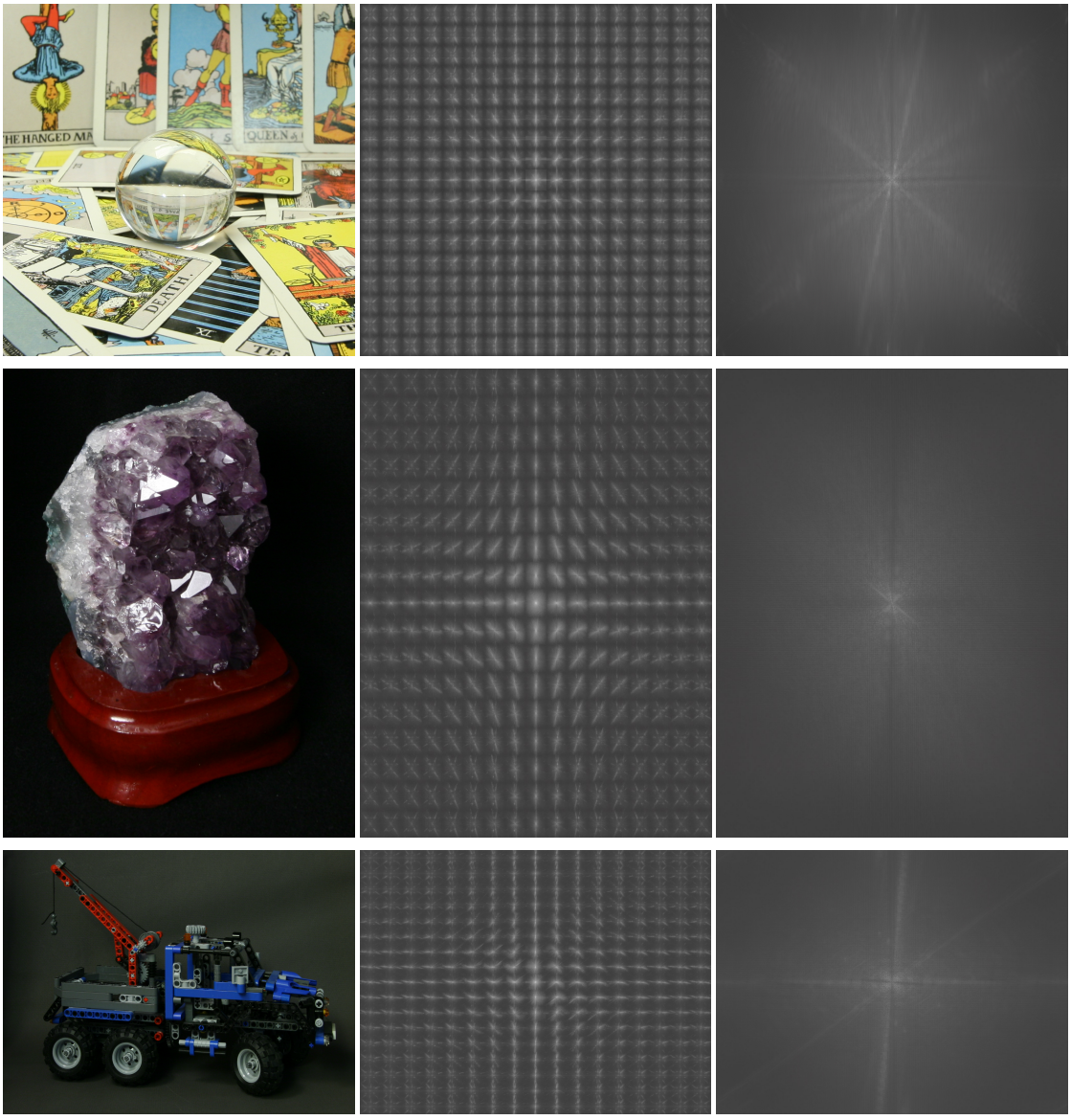
One of the conclusion of the work in [17] is that the non-zero part of the light field spectrum lies in a 3D subspace of its 4D Fourier space. This property is known as *dimensionality gap*. A representative work for light field photography includes reconstruction from coded aperture in [19]. With no depth information, Levin *et al.* [20] use only 1D trajectory of angular samples to reconstruct Lambertian views.

Another sparsity structure was proposed in a recent work in [21]: it assumes that a light field can sparsely be represented with a training-based dictionary, *i.e.* that it has a similar structure to the pre-trained light field data.

Moreover, image analysis in the Fourier domain has been a interesting tool in computational photography and rendering algorithms. In fact, the sparsity of natural signals spectra, such as light fields, makes it possible to reconstruct them from a small set of samples [19, 20]. This sparsity derives from the continuous Fourier transform, where continuous-valued depth in a scene translates to 2D sub-spaces in the Fourier domain [17]. In this context, Shi *et al.* [22] proposed to reconstruct a full light field from a 1D trajectory of selected views, leveraging the sparsity of the light field in the continuous 4D Fourier domain. The level of sparsity is constrained here by the number of input viewpoint samples. Non-Lambertian light field images can be reconstructed in this case, with specularities being conserved. The observation that much of the sparsity in continuous spectra is lost in the discrete domain has lead to an optimization step of sparsity in the continuous domain [22] that remarkably improves the reconstruction quality.

To illustrate this sparsity, Figure 1.6 shows the spectrum of different real light fields from the Stanford dataset [6]: the spectrum of each light field is represented in two different ways, using the order of respectively spatial and angular coordinates, specified in the second row illustrative schemes.

[6] http://lightfield.stanford.edu/lfs.html



Central sub-aperture image | from sub-aperture images | from the lenslet image

Figure 1.6: Visualization of the 4D Fourier spectrum of the light fields *Crystal*, *Amethyst* and *Lego Truck*.

The works in this thesis have built on this sparsity property of the Fourier transform of light fields, to propose solutions to compression and compressive reconstruction of this type of content.

1.5 Applications

Among the application examples that light field content permit, we can distinguish depth estimation that extracts the geometry from light fields, and image rendering methods for generating new images. Other applications not directly related to these two categories can use the depth/geometry information to study the scene and extract more features about it. We propose in the following a brief description of these different applications.

1.5.1 Depth Estimation

The 4D light field representation essentially contains multiple viewpoints of the scene, which makes depth map estimation possible. Unlike stereo-based techniques, light field-based depth extraction does not require camera calibration, which makes it more suitable for real-world data acquisition. Furthermore, the light field structure expands the disparity into a continuous space [23]: this property appears mainly in epipolar plane images (EPIs) for densely sampled light fields, where corresponding pixels are projected along a line. Existing depth approaches can be divided into three main categories: EPI-based approaches, sub-aperture image matching-based methods, and learning-based methods.

In epipolar images, the slopes of the lines indicate the depths of objects [6]. So, most EPI-based methods for depth estimation rely on estimating the slopes in EPIs using various optimization algorithms. Using a structure tensor in EPI space, the local direction of each line can be estimated [23] and the corresponding depth can be deduced. The estimated depth is refined by building certainty maps using structure information in [24]. Applying a confidence measure in EPI space, the reliability of estimated depth is further computed to take into account occlusions and generate the final depth map [25].

The small baseline of most plenoptic data makes the disparity range of sub-aperture images quite narrow. To match sub-aperture images with extremely narrow baseline, a cost volume is computed in [26] using the similarity measurement between sub-aperture images and the central sub-aperture image shifted at different sub-pixel locations. Occlusion-aware algorithms for depth estimation identify occluded edges by ensuring photo-consistency in only occluded regions [27] and can handle multi-occluder occlusion by using the concept of surface camera [28, 29].

Among other cues that can be exploited in depth estimation, the defocus cue is the prominent one. Ng's work [14] explains how a light field can be refocused on any region by shearing it with both a digital and a Fourier-based refocusing, at different candidate depths. Based on this work [14], the defocus cue and the correspondence cue are combined to estimate the depth [30]. An additional shading cue refines details in the shape [31]. Two other cues, angular entropy metric and adaptive defocus response [32], are proposed to improve the robustness of depth maps to occlusions and noise.

More recently, new techniques based on Convolutional Neural Networks (CNNs) have been proposed for depth map estimation, either by learning a mapping between the 4D light field and a representation of the depth map in 2D hyper-plane orientations [33], or sparse decomposition [34] to relate the depth to the orientation in EPIs.

1.5.2 Light Field Super-Resolution

Due to the limited sensor resolution and processing speed in plenoptic cameras, many researches have been proposed to super-resolve captured light fields in different dimensions. We investigate here light field super-resolution mainly in spatial and angular dimensions. Nevertheless, some approaches have also focused on super-resolving the spectral dimension [35, 36].

Spatial Super-Resolution Even if a straightforward way to spatially super-resolve a light field is to apply known super-resolution techniques on each view individually, this means under-using the information that a light field exhibit, especially in terms of inter-view similarities. The work in [37] takes advantage of the non-integral shifts between corresponding pixels in two views, to propagate pixel intensities to a target view from its neighboring views. Using the same principle, various methods estimate the depth to derive these non-integral shifts [23,38,39]. Other approaches use hybrid systems to super-resolve light field images based on a single high resolution image captured via a DSLR camera [40], with an additional depth-based view synthesis to better recover high frequency information in low resolution images [41]. CNN-based methods [42, 43] perform spatial super-resolution of light fields without depth estimation.

Angular Super-Resolution Many works have focused on angular super-resolution for light field using only a small set of views, that can also be seen as a reconstruction of the plenoptic function from a sample subset. A first category of these works start by estimating depth to warp the available view to new views, by computing the optical flow between neighboring views [44], or using a super-pixel segmentation [45] or optimizing a cost function that ensures a robustness to errors in the estimated depth [46]. Since depth map estimation can be much sensitive to occlusions and textures in light field images, the quality of synthesized views by these approaches can be heavily impacted. With the recent advances in research in deep learning methods, some CNN architectures have been proposed to provide a better quality of novel synthetic views [47–49]. As depth-based view synthesis approaches tend to fail in recovering occluded regions, or specularities, alternative approaches are based on sampling and reconstructing the full-resolution light field. As direct interpolation solutions may lead to aliasing effects in the resulting views, some works have explored light field reconstruction in the Fourier domain [20, 22]. However, a general limitation of these approaches is that they require a specific sampling pattern. Instead, Vagharshakyan *et al.* [50, 51] considered angular super-resolution as an inpainting problem on the epipolar palne images, and used an adapted discrete shearlet transform to super-resolve the angular resolution of a light field.

1.5.3 Rendering and Refocusing

A 4D light field can be seen as a set of views captured by cameras at positions defined by the two parallel planes. The planar representation samples the light rays according to the uv and xy coordinates. In practice, with a sufficiently dense set of cameras, one can virtually render the light field at any position of the plane of cameras, or even in another plane closer to the object, by re-sampling and interpolating the light rays [52] from the existing views (using the fact that the radiance of the rays remains constant in free space), rather than using geometry information to synthesize the view [53].

Figure 1.7: Refocusing of the light field *Crystal* on two different planes.

However, in light field rendering, aliasing effects may occur in the new synthesized views if one does not have a sufficient number of samples. Nevertheless, acquiring too many samples of a light field is not practically feasible. Investigations about the minimum number of samples required for light field rendering from a plenoptic acquisition leaded to the conclusion that the maximum disparity between neighboring views should be less than one pixel (*i.e.* the camera resolution) to render new views without generating aliasing [54, 55]. If the geometry of the scene is known, the number of samples needed to accurately render the light field can be reduced. A generalized model combining light field rendering with Depth Image-Based Rendering techniques (DIBR) was proposed in [56] to render novel views from unstructured input, *i.e.* samples at irregular positions.

Furthermore, another very largely known application of light field imaging is refocusing. Indeed, with the geometrical information that a light field provides, it is possible to retrieve small details in the scene while generating images at different focal planes. Digital refocusing is performed by shearing the 4D light field and then summing over the angular dimensions to synthesize refocused images [14]. A refocusing example is given in Figure 1.7. Moreover, with a known depth information, one can generate images refocused in all depth plans of the captured scene, forming the focal stack. An all-in-focus image can then be synthesized by choosing, for each pixel position, its value in the refocused image where it is in focus (see Figure 1.8).

Besides, another aspect of light fields is its ability to show objects through occlusions by synthesizing large aperture. Indeed, since the occluder and the object of interest are at different layers of depth, light field views can capture different small portions of the object [57, 58]. A virtual aperture image can then be synthesized by propagating the area of interest at a specific depth layer.

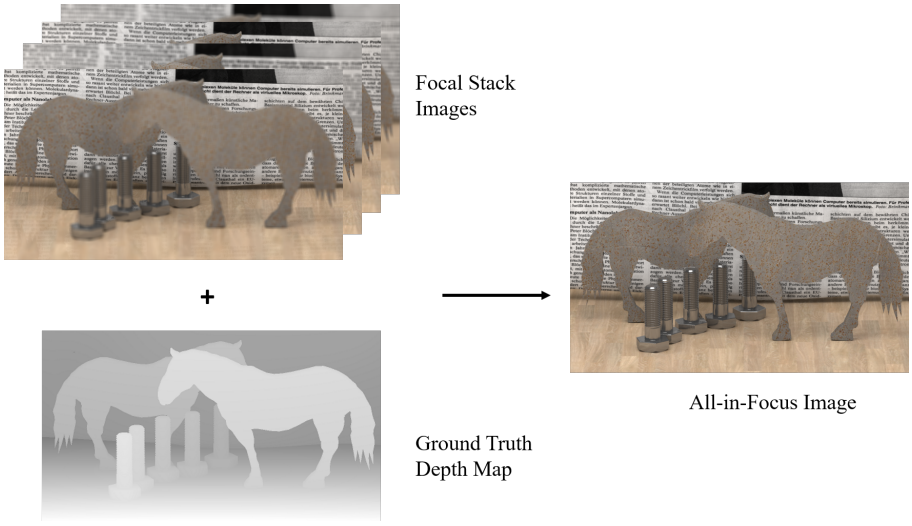


Figure 1.8: Generating the all-in-focus image using the focus stack and the ground truth depth map: each pixel value is chosen from the focal stack in which the pixel is in focus.

1.5.4 Further Applications

The geometry information that light fields contain can also be exploited for many other purposes than image rendering and depth map calculation.

Indeed, the correlation in light field views can help ensuring label consistency in segmentation approaches [59–61]. Lights fields have been used to segment refractive surfaces and transparent objects into superimposed layers [62, 63].

Moreover, with their inherent robustness to occlusions, light fields have proven a good performance in the tasks of object detection and classification [64–66]. Compared to object recognition, material recognition is a more challenging problem because of the diversity in the appearance of materials that may depend on lighting conditions.

Using a recent learning-based techniques, a CNN-based material recognition approach was proposed in [67] using a 4D light field dataset. The best performing CNN architecture achieved a 7% boost compared to conventional approaches based on single 2D image.

Chapter 2

Compression and Compressive Photography

Light field compression is a critical field of research for the practical usage of this type of data. Indeed, light fields involve a large volume of data that needs to be efficiently acquired, stored and transmitted. Light fields also capture scenes from different viewpoints, and thus exhibit redundancy in both spatial and angular dimensions: these redundancies can be observed between sub-aperture images or in sub-regions of the lenslet image (Figure 1.2).

Compression approaches adapted to light field content are indispensable to offer efficiency in light field storage, transmission and display. Another way of tackling the problem of the large data amount in light fields is to propose solutions for acquiring less data and restoring the whole light field offline. This section provides an introduction to the background related to light field compression and compressive acquisition.

2.1 Light Field Image Compression

Despite the significant progress in light field acquisition systems, rendering new images in real time is still a challenging task due to the computational limitations and the high data rate constraints. Fortunately, light fields exhibit data redundancy in both the spatial and angular dimensions [52], which can be effectively removed using compression techniques.
Several compression approaches have been proposed for light field content and can be divided into three main categories: transform coding, predictive coding and pseudo-sequence-based coding. Some methods also use hybrid coding schemes.

Moreover, some research related events, such as the "ICME 2016 Grand Challenge on light field image compression" [1], were held to explore new efficient compression methods for light fields. The dataset proposed for the test is composed of real images acquired using a plenoptic camera. Current state-of-the-art techniques can achieve more than 60% bit-rate reduction [68] compared to HEVC intra coding for lossy compression of plenoptic images. However, with the progress of light field acquisition, and the increasing need to capture light fields with much larger

[1] https://mmspg.epfl.ch/meetings/page-71686-en-html/icme2016grandchallenge_1

baselines and higher spatial resolution than plenoptic images, better performance is expected for light field compression schemes, along with reasonable resource requirements in terms of data rates, computational complexity, and power consumption.

Furthermore, the compression process has a certain distortion impact on the structure of light fields, especially with large disparities, and can influence the subsequent processing such as refocusing [69], depth estimation or super-resolution. Thus, obtaining a decent bit-rate reduction while preserving the light field structure is also a challenge for compression solutions. In this section, we introduce quality assessment approaches and propose a detailed overview of the state-of-the-art approaches for light field image compression.

2.1.1 Quality Assessment

Light field quality assessment presents a important aspect that helps to obtain a better understanding of the performance of light field acquisition, generation, and processing techniques. It involves both spatial (pixel) quality and angular (view) consistency. Fu *et al.* [70] showed that a light field camera presents a more stabilized visual resolution in terms of depth of field (DOF), compared to a 2D classic camera, thanks to the refocusing possibility. Regarding the spatial quality evaluation, Peak Signal to Noise (PSNR) and Structure Similarity index (SSIM) [71] are very often used [72]. Besides, several applications such as angular super-resolution and cross-view light field editing require to measure the inter-view coherence. This assessment is usually calculated using the epipolar plane image (EPI) representation of the light field. Indeed, Adhikarla *et al.* [73] introduced a light field viewing setup to perform subjective evaluation of the angular coherence. They proposed furthermore an adapted SSIM measure to light field angular assessment.

2.1.2 State of the Art of Light Field Image Compression

Regarding light field coding, several solutions have already been explored using either simple tools such as Vector Quantization (VQ) followed by Lempel-Ziv (LZ) entropy coding [52], or other signal processing-based methods. In the Vector Quantization approach [52], light field images are partitioned into blocks represented as vectors, and a small subset of the vectors is trained to approximate the entire vector space.

The early light field coding schemes adopted transform-based approaches using Discrete Cosine Transform (DCT) or Discrete Wavelet Transform (DWT). Classical coding approaches like JPEG (using DCT) or JPEG2000 (using DWT) were applied to light field raw image compression, but did not achieve sufficient quality results since they are not designed to specifically deal with light field data.
In [74], a 3D DCT was proposed to exploit cross-correlation between the different views of a light field. A selected subset of views and its neighbors were arranged into a 3D structure, and a 3D DCT was used to generate the de-correlated sub-views group. Recently, a 4D DCT-based light field codec was proposed in [75] aiming at exploiting the whole 4D redundancy: the light field is partitioned into 4D blocks, and the coefficients resulting from their 4D DCT transforms are grouped generating a stream encoded using an adaptive arithmetic coder. Besides the DCT, DWT was also used to exploit the spatial and angular coherence in the light field views [76, 77].

In [78], the LF content was separated into various sub-aperture images by extracting one pixel with the same position from each micro-image and a 3D DWT was then applied to a brick of the sub-aperture images. A 4D DWT was presented in [79] to directly compress the light field without using view arrangement.

These transform-based approaches may create misalignment when coding light fields with large inter-view disparities. To overcome this limitation, a disparity compensation was incorporated in [80] into the lifting structure for the DWT across the sub-aperture images to exploit their correlations, along with another shape-adaptive DWT applied to encode the resulting coefficient images. Dong *et al.* [81] proposed a hierarchical disparity compensated coding scheme, where they first decomposed a light field into sub-bands using wavelet packet transform. The wavelet packet bases were divided into two groups: predictable bases and unpredictable bases. A disparity map was then applied for sub-band prediction.

An alternative way to compress light field images was proposed in [82], where the focal stack (which consists of images focused at different depth values) is encoded using 3D DWT and SPIHT scheme. The entire light field is then reconstructed from the decompressed focal stack using a combination of dimension reduction and a 2D filtering. Moreover, a hybrid compression scheme was introduced in [83] where a 2D DWT is applied to each micro-image followed by a 2D DCT applied to sets of DWT coefficients from neighboring micro-images.

Besides, Principal Component Analysis (PCA) was proposed in [84] to compress light field images, by ordering the sub-aperture images into column vectors of a matrix, applying PCA locally/globally. The dominant eigen vectors are then separately encoded with an image encoder.

Moreover, with the improvement of standard video codecs, the recent light field coding approaches considered using the latest HEVC [85] video compression standard, taking advantage of its novel prediction tools. Some methods compressed directly raw light fields as a single image, and employed the intra coding mode of the standard video encoder. The light field image corresponded to a 2D array of micro-images presenting a cross-correlation in a neighborhood.

Conti *et al.* introduced in [86] the concept of self-similarity into HEVC for new prediction modes adapted to light field compression. Similar to motion estimation, a block-based matching algorithm estimates the best predictor block for the current block over the previously coded and reconstructed part of the current picture. Furthermore, the bi-directional self-similarity compensated prediction and estimation were proposed in [87] to improve compression efficiency. Monteiro *et al.* [88] proposed another prediction scheme where a set of similar blocks are chosen and linearly combined to predict the current block more accurately.

More recently, Liu *et al.* [89] introduced a Gaussian Process Regression-based prediction scheme. The prediction was here modeled as a Gaussian process from a set of nearest neighbor blocks to form the predictor block. Li *et al.* [90] compressed lenslet images from focused plenoptic cameras by directly applying a displacement intra prediction scheme into HEVC to better explore the redundancies in light field images.

Some other coding schemes proposed instead to extract the sub-aperture images from the light field content to represent it as a multi-view content. The multi-view-based coding of light fields used 3D video coding solutions such as Multi-view Video Coding (MVC) [91] (in [92,93]) or

the multi-view extension of HEVC (MV-HEVC) [94] (in [95]). Another version of the multi-view HEVC extension for light field video was introduced in [96], where the inter-view prediction was represented in a two-directional parallel structure.

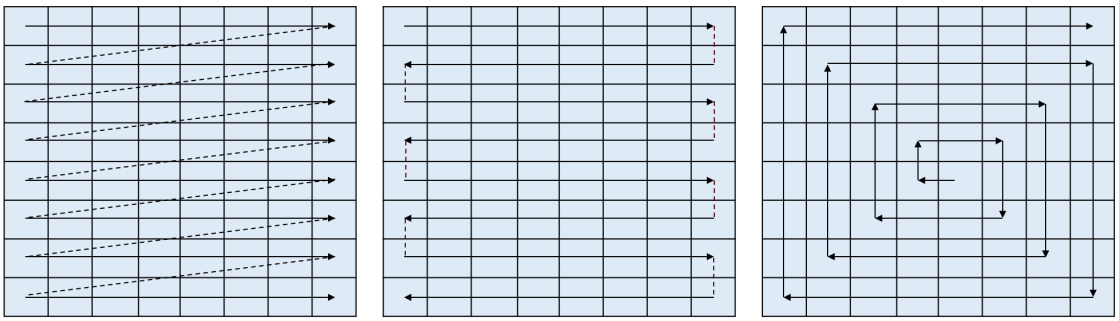


Figure 2.1: Typical rearrangement paths for pseudo-sequence-based coding approaches. Left to Right: zig-zag, raster and spiral.

Furthermore, pseudo-sequence-based approaches encode the light field viewpoints using a 2D video coding standard such as H.264/AVC [97] (in [98]) or HEVC [85] in [99]. These methods re-organize the sub-aperture images of a light field as a pseudo-sequence to be encoded with standard video encoders, and can achieve good compression performances. In [100], Viola *et al.* reported that the self-similarity-based light field image compression methods cannot achieve the same performance as pseudo-sequence-based methods due to their inflexibility to exploit correlations between sub-aperture images, especially in low bit-rate.

Taking account of high correlations that adjacent viewpoints can have, various scan orders of the set of sub-aperture images and prediction structures were proposed to explore the use of the video encoders in light field compression: raster, zig-zag, spiral, lozenge, U-shape, hybrid [99, 101–104] (see examples in Figure 2.1). Liu *et al.* [89] proposed a spiral scan of the viewpoints, where the central sub-aperture image is set to I-image, and the rest of viewpoints are set to P- or B-frames in symmetric 2D hierarchical structure. Different QPs were assigned to frames using an empirical bit allocation scheme. Li *et al.* [105] partitioned each sub-aperture image into four quadrants, in which the sub-aperture images were encoded hierarchically in both horizontal and vertical directions, the reference frames were chosen based on the distance between views. An optimal bit allocation algorithm was further added to exploit the viewpoint correlations, taking into account the influence of different viewpoints on the following viewpoint coding.

Perra and Assuncao [106] divided the raw lenslet image into multiple tiles of equal size, and organized them into a pseudo-video sequence, encoded using the HEVC *inter* prediction to exploit correlations among frames. To find an optimal scan order, a homography-based low-rank approximation was presented in [107]; this method can adaptively adjust the scan order according to the light field content. Zhang *et al.* [108] proposed a disparity correlation-based prediction method to encode light field sub-aperture images: the images are ordered following a rotation scan starting from the central view, and sent in a sequence to the HEVC encoder. Recently, Ahmad *et al.* [109] proposed an adaptive prediction and rate allocation scheme to efficiently

compress light field images in a pseudo multi-view video using the multi-view extension of high efficiency video coding (MV HEVC) [94].

Another way to exploit inter-view similarities in light field contents is to use a reduced amount of the data to encode, to remove the redundancy before encoding, and reconstruct the integral light field data. This can be integrated into a scalable coding scheme in which a base layer encodes a sampled set of the light field content and one or more layers to reconstruct and decode the full image. Dricot *et al.* [110] used a limited number of views from an original integral image to reconstruct an integral image. The residual obtained from the reconstructed integral image and the original one, as well as the extracted views are encoded into bitstreams using respectively HEVC and 3D-HEVC.

Conti et al. [111] proposed a scalable 3D holoscopic video coding scheme to support display scalability, in which views are separated and encoded in multiple layers (one base layer and two enhancement layers), and an inter-layer prediction method was proposed to improve the coding efficiency.

In [68], a sparse set of micro-lens images (also called elemental images) was encoded in a base layer. The other elemental images were reconstructed at the decoder using disparity-based interpolation and inpainting. The reconstructed images were then used to predict the entire plenoptic image and a prediction residue was transmitted yielding a 3-layer scalable coding scheme. However, this approach is only suitable for focused plenoptic images.

2.2 Light Field Compressive Photography

With the emergence of massive amounts of high-dimensional data such as light fields in different domains, the need for efficient capturing of this type of visual data has became a widely interesting research area. Indeed, computational photography solutions come to provide a richer visual perception, and capture more information, which makes the recorded scene more flexibly used for various applications (*e.g.* immersive applications). While this can involve multiple cameras and sensors, the amount of data produced is often higher than the capacities of the existing storage devices. Besides, real-time compression of such multi-dimensional data can be very costly.

In this section, we present an overview of the state-of-the-art solutions related to light field compressive acquisition. There are several ways to tackle the problem of compressively capturing the scene information and then reconstructing the light field. Each category is detailed below.

2.2.1 Coded Aperture Imaging

Numerous efforts have already been presented to improve the spatio-angular resolution trade-off of acquired light fields. Liang *et al.* [112] proposed a programmable aperture approach which exploits the fast multiple-exposure feature of digital sensors to sequentially capture multiple subsets of light rays, but at the cost of a longer exposure time. In [113], two attenuation masks are used, one placed at the aperture and the other one in front of the 2D photo sensor. Zhang *et al.* [114] presented a phase-based approach to reconstruct a 4D light field using a micro-baseline

stereo pair. Yagi *et al.* [115] proposed an aperture pattern and reconstruction algorithm derived via Principal Component Analysis. In this method, the compressive pattern chosen is dependent to the signal, which presents a limitation since the original signal may not be known depending on the application. Another approach was proposed later in [116] where the basis vectors are derived from non-negative matrix factorization (NMF).

2.2.2 Light Field View Synthesis

The problem of providing higher resolution light fields can be seen as an angular super-resolution problem, in which additional views are synthesized from a few number of views.

In this context, some recent methods aimed to generate a full light field from a light field view subset. Wanner *et al.* [23] introduced a variational light field angular super-resolution framework by utilizing the estimated depth map from the input views to warp them to novel views. Several works also proposed light field synthesis using deep learning and Convolutional Neural Networks (CNN). In [48], a full light field is generated from its four corner views, using two sequential CNNs, one for disparity estimation then another for color prediction on the top of warped images. With a similar pipeline, Srinivasan *et al.* [49] addressed the problem of extrapolating a light field from a single central view, with good visual results. Note that both those works only considered plenoptic contents that present limited parallax.

2.2.3 Compressive Sensing-based Acquisition

One possible alternative to overcome the problem of costly storage and compression is to directly acquire a compressed data. This idea is related to the *Compressed Sensing* theory [117]. Several contributions have been proposed in this new field of signal processing, and theoretical demonstrations have opened the possibilities to introduce new solutions for multi-dimensional signal acquisition and reconstruction.

We introduce in the following paragraph the theoretical support related to the compressed sensing that provides the guarantee for compressive acquisition and reconstruction of light field data.

Compressed Sensing Theory

In the context of signal processing, the Nyquist-Shannon theorem for sampling band-limited continuous-time signals presented a fundamental component in designing acquisition systems. It states that if a signal contains no frequencies higher than a certain f, it can completely be recovered by giving a series of samples spaced at a rate larger than $2f$, called the Nyquist rate. In practice, however, designing systems that operate at the Nyquist rate is challenging despite the continuous development of computational capacities [118]. Indeed, the required Nyquist rate is so high that the capture of many samples is needed, which may be too costly or physically infeasible for certain applications.

In this context, the emerging Compressive Sensing (CS) theory has gained great attentions and became an interesting research field, used in many applications in signal processing and computer science. Compressive sensing suggests a new framework for sampling and compressing data. Moreover, it offers alternative acquisition systems in which a compressive representation

of the signal is directly acquired with only a small subset of measurements.

In addition, it has proved that high-dimensional data can be accurately recovered from sampled low-dimensional entries, as long as the *sparsity* [119] condition is satisfied.

Sparsity describes a signal as a linear combination of a few vectors of a dictionary (or basis). To formulate the last statement, let $x \in \mathbb{R}^N$ be a signal, $\Psi \in \mathbb{R}^{N \times N}$ a basis for $\mathbb{R}$, where columns are basis vectors. We can write:

$$x = \sum_{i=0}^{k} \alpha_i \psi_i, \tag{2.1}$$

where α_i are the coefficients of x in Ψ and $\psi_i \in \mathbb{R}^N$ are elements in the basis Ψ.

A vectorized formulation of Eq. 2.1 is:

$$x = \Psi\alpha, \tag{2.2}$$

where α contains all α_i. We say that x is k-sparse in the basis Ψ if $k << N$.

Besides, the Compressive Sensing (CS) theory suggests to approximate a signal from compressive representation, composed of a small number of measurements. Indeed, as long as the basis is chosen to fit well the structure of the signal x, an accurate approximation can be obtained by sparse representations with $k << N$. Since only few elements are used, compression is a direct byproduct of sparse representations. The quality of the sparse representation (*i.e.* the error level introduced by the model) is then directly inherent to the chosen basis. Actually, the more sparse the representation is, the higher the compression ratio can be. In the case of image signals, Discrete Cosine Transform (DCT), Discrete Wavelet Transform (DWT) and Fourier Transform (FT) are good candidates for the basis Ψ.

Let M be the number of randomly chosen vectors $\phi \in \mathbb{R}^N$. $\Phi = [\phi_1, \phi_2, ..., \phi_M]^T$ is named measurement matrix. The CS sensor produces the measurements $y = \Phi x \in \mathbb{R}^M$. Assuming that x is k-sparse in the basis Ψ, we have $x = \Psi\alpha$, where $||\alpha||_0 \leq k$. In this setup, reconstructing the sparse signal x from the measurements in y involves solving the following optimization problem:

$$min||\alpha||_1 \quad s.t. \quad y = \Phi\Psi\alpha$$

and then computing the estimate $\hat{x} = \Psi\hat{\alpha}$.

Several works [3,120,121] demonstrated that a signal of size N can be accurately recovered at $s << N$ non-zero elements using Gaussian sensing matrices, in the condition that $k \geq Csln(N/s)$, where C is a universal constant. Hence, the number of necessary samples s depends on the sparsity k of the signal and its length N. The compressed sensing then makes it possible to exactly recover sparse signals using a significantly reduced number of measurements, compared to the Nyquist rate.

Compressive Sensing for Light Fields

Several research efforts have been recently dedicated to explore light field compressive acquisition schemes, in order to overcome the existing light field sensing limitations. The authors in [122] proposed two separate compressive acquisition architectures, one exploiting spatial correlations

and another exploiting angular correlations in coded lenslet images.

Similarly, a random-coded mask is used in [123] to capture random linear combinations of angular samples, and the light field is rebuilt via a hierarchical Bayesian framework. Wang *et al.* [124] further improve the quality of reconstructed light fields using a random convolution CMOS sensor, which is able to maintain more information by means of correlation.

An architecture using coded projections on the sensor image is described in [21], where the light field is reconstructed using sparse methods with a learned over-complete dictionary.
Recently, Miandji *et al.* [125] proposed a framework that trains an ensemble of dictionaries operating locally on a set of patches extracted from training data. A sub-sampled input data is then reconstructed using a reduced union of sub-spaces model, assumed to represent signals more sparsely than classical dictionary-based methods. However, the method is limited by the size of the dictionary vectors that depends on the training set: *i.e.* to reconstruct a new differently-sized light field, it requires to retrain dictionaries with light fields of the same size. A multi-shot sensing system is proposed in [126], where different color coded masks are used in multiple shots and the light field is reconstructed by a training-based dictionary in a compressive sensing framework.

Gupta *et al.* [127] trained a two-branch network to decompress compressed light field for various sensing frameworks. One of those branches is a fully connected network which limits the patch size.

Vadathya *et al.* [128] also designed a deep learning-based approach. From the compressed image, they extract the central view and a disparity map, which are used to reconstruct the final light field by warping the central view. For that purpose, they use three different CNNs. Note that the network architectures in both [127] and [128] are adjusted for only one mask pattern. Therefore, they are not invariant to different locations on the sensor as each patch on the sensor is generated by a different compressing matrix. Last, Nabati *et al.* [129] presented a CNN-based system to synthesize a light field from a color coded image and its correspondent coding matrix.

2.2.4 Sparse Fourier-based Reconstruction

As previously explained, sparse reconstruction algorithms can rely on various representations to approximate a signal from a subset of measurements, as long as the chosen representation can efficiently describe the structure of the considered signal.

In their work based on Sparse Fourier Transform [130] [131], Shi *et al.* highlighted the sparsity of light fields in the Fourier domain, especially in the angular dimensions [22], which corresponds to the redundancy views exhibit with each other. Shi *et al.* recover the light field signal from an angular subset of views that are not sub-sampled spatially. The subset choice in [22] is limited by a structure that can ensure a good initialization of the frequency estimation.

Furthermore, an optimization step comes to approximate the spectrum to the continuous Fourier domain by refining the frequency positions with a small non-integer step, the sparsity being better preserved in the continuous than in the discrete domain. Indeed, in Figure 2.2, we can see that in the discrete Fourier domain, the spectrum of the light field slice presents *sinc* (cardinal sine) tails and thus is not sparse, while in the continuous domain the Fourier spectrum is much sparser, and consists of four peaks that do not fall on the grid points of discrete Fourier

transform.

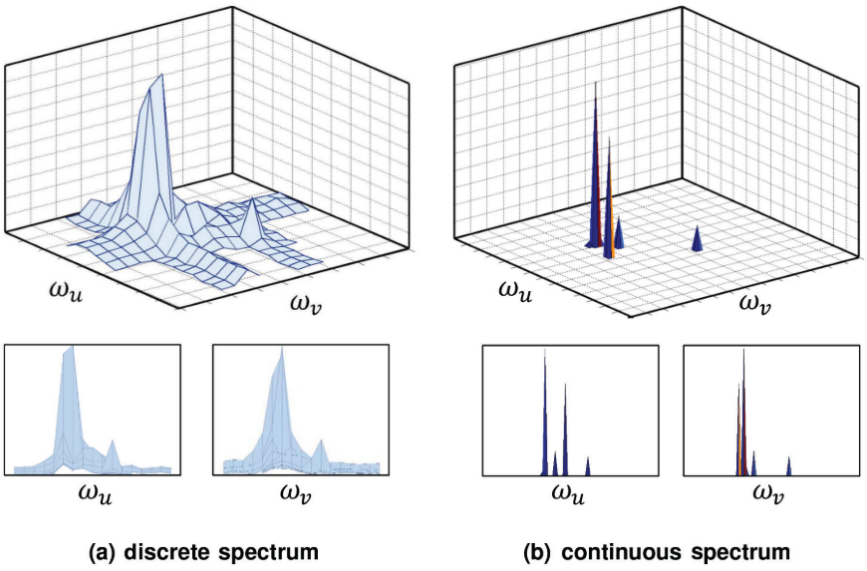


Figure 2.2: Representation of the Fourier transform of 2D slice of the light field *Crystal* from the Stanford dataset.

Another interesting algorithm for the sparse signal reconstruction is the Frequency Selective Reconstruction (FSR) introduced in [1]: the objective is to re-sample 2D image from a non-regularly sampled subset of pixels. The reconstruction is conducted in the Fourier domain, and consists of an iterative selection of frequencies that best fit the available samples of the signal and reduce the residual error until a sufficient recovery quality is achieved. The original work that introduced this idea of frequency selection for signal recovery idea was proposed in [132]: it aims to iteratively approximate an image content by a weighted linear combination of basis functions to extrapolate the missing blocks in the image. The extrapolation is conducted in the Fourier domain, since 2D Fourier Transform are suited for the problem of signal extrapolation. The data loss is supposed here to be in regularly spaced blocks of the image, where a full neighborhood area is present. Improvements to the Frequency Selective Extrapolation algorithm were introduced, either by adding a spatial weighting function [133] to the error criterion, or defining a deficiency compensation parameter [134] to overcome the non-orthogonality of frequencies defined in the error block space. A *best approximation* was used in [135] for spatio-temporal prediction (originally introduced in [136]). Compared to Frequency Selective Extrapolation where the residue is updated just by the selected basis function in each iteration step, the *best approximation* method modifies the expansion coefficients of all the already selected basis functions in order to produce the best possible approximation using the selected set. Other works proposed further algorithmic improvements such as motion compensation [137,138], processing order optimization [139,140] or texture-dependent reconstruction [141]. Besides, there were several applications of the Frequency Selective Extrapolation such as video coding [137, 142], multi-view data reconstruction [143] or mesh-to-grid image resampling [144].

These notions are used throughout this thesis by exploiting light field sparsity or compressibility to tackle the problems of light field image acquisition and coding. In particular, we propose solutions utilizing sparse models for efficient light field compression and compressive acquisition. Enhancing the sparsity of the representation is also a key element that has been explored in our contributions.

Part II

Contributions

Chapter 3

A Novel Scalable Scheme for Light Field Image Compression

3.1 Introduction

With the ever-growing interest in light field imaging, capturing higher scale light field data has become more and more required. The acquisition of spatially and angularly highly-resolved light fields yields a huge amount of data, hence, it poses real challenges in terms of storage and transmission requirements. Thus the need to introduce more efficient compression schemes for light fields.

In fact, light fields involve a large volume of data that needs to be efficiently acquired, stored and transmitted. However, light fields capture scenes from different viewpoints, and thus exhibit redundancy in both spatial and angular dimensions. These redundancies can be observed between sub-aperture images or in sub-regions of the lenslet image, and can be exploited in designing compression schemes of light field data.

We present in the following a novel scalable compression scheme for light field images. The input data is a sparse subset of light field sub-aperture images, encoded using a standard video codec and used to reconstruct the whole light field in the Fourier domain. The proposed method is efficient, does not require any prior knowledge of the scene geometry and can be utilized for different light field contents, sparse or dense, and even with large baselines.

3.2 Proposed Light Field Compression Scheme

We introduce here a scalable coding approach for light field images using a sparse set of sub-aperture images and a sparse Fourier-based reconstruction method. The goal is to reduce the cost of light field coding by transmitting only some given viewpoints, and reconstructing the missing ones by exploiting the redundancies that light field views exhibit. We assume that these redundancies yield a sparsity property in the angular (view) domain.

We select a set of light field sub-aperture images (or views) and encode them in a first layer in a pseudo-sequence structure using the standard encoder HEVC. Once these views are decoded, a full light field is reconstructed using the sparse Fourier transform-based method in [22] in the 4D continuous Fourier domain. A quality refinement of the resulting light field is then obtained via a second layer where the residual error to the original light field is encoded. The proposed approach is a two-layer scalable structure. From the first to the second layer, quality scalability is enabled.

Let $L(x, y, u, v)$ denote the 4D representation of a light field, describing the radiance of a light ray parameterized by its intersection with two parallel planes [5]. The angular (view) coordinates are denoted with (u, v), where $u = 1 \dots N_a$ and $v = 1 \dots N_a$. The spatial (pixel) coordinates are denoted with (x, y), where $x = 1 \dots N_x$ and $y = 1 \dots N_y$). The view at the angular position (u, v) is defined as $I_{u,v}$.

The main steps of the proposed compression scheme are depicted in Figure 3.1. First, a sparse set $\{I_{\mathbf{p}}\}_{\mathbf{p}\in P}$ of light field views at pre-defined positions in P is selected (the selection pattern is shown in blue in Figure 3.2), and encoded as a video sequence using HEVC [85] in a base layer (BL).
Then, the set of non-selected views $\{I_{\mathbf{q}}\}_{\mathbf{q}\in Q}$ is reconstructed using the decoded views $\{\hat{I}_{\mathbf{p}}\}_{\mathbf{p}\in P}$ ($Q \cup P = \Omega$, Ω denoting the entire set of view positions). We use a reconstruction method that exploits the light field signal sparsity in the angular Fourier domain [22]: Sparse Fast Fourier Transform (SFFT).
Finally, the restored light field is used as an inter-layer predictor of the original light field, in an enhancement layer (EL) using SHVC [145], leading to a two-layer SNR-scalable scheme.

3.2.1 Base Layer Coding

A subset of light field views $\{I_{\mathbf{p}}\}_{\mathbf{p}\in P}$ is first selected and then encoded in HEVC as a YUV sequence (see Figure 3.2), delivering the base layer bitstream. The order in which the viewpoints are placed in the sequence ensures to have the highest correlations between successive frames, as shown by a red line arrow in Figure 3.2. The corresponding decoded views $\{\hat{I}_{\mathbf{p}}\}_{\mathbf{p}\in P}$ are later used to reconstruct the remaining views.

Light Field Viewpoint Reconstruction

In [22], Shi *et al.* leverage the fact that the 4D Fourier transform of a light field is sparse, and use the Sparse Fourier Transform [130] to compute the Fourier transform and invert it to recover the full light field data array. The inter-view redundancies make it possible to represent sparsely the light field data in the Fourier domain, especially in the angular domain. The methods uses this sparsity to reconstruct the full light field from a subset of sub-aperture views.

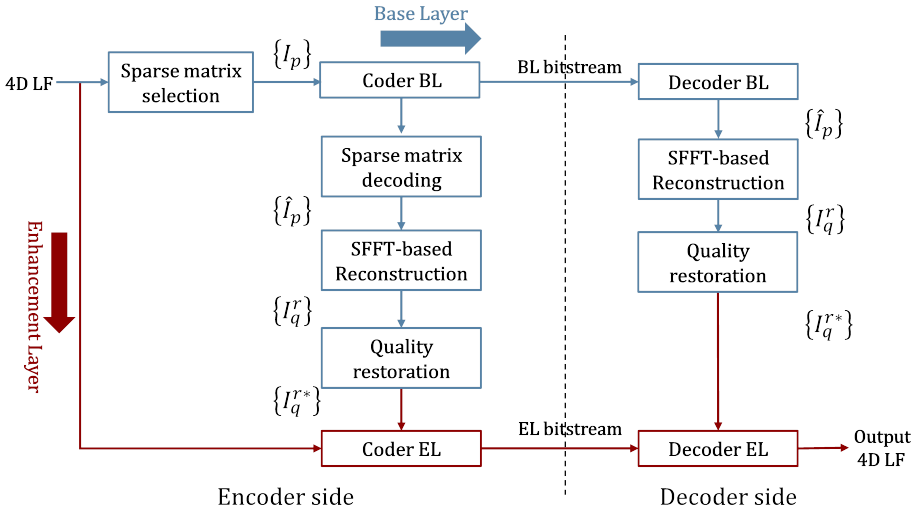


Figure 3.1: An overview of the proposed compression scheme.

In contrast to the prior works where a specific sparsity structure assumption is made, the work in [22] uses a 1D path of angular samples, and reconstruct the light field at the sparsity level constrained by the number of selected samples.

Besides, Shi *et al.* [22] explain the advantage of the reconstruction in the continuous Fourier domain, compared to the discrete Fourier transform where the sparsity can be reduced if the discrete sampling grid does not coincide with the spectrum peaks, which is known as the windowing effect. Indeed, sampling a signal, such as a light field, inside some finite window is equivalent to multiplying it by a box function (or a convolution by an infinite *sinc* in the frequency domain). If the non-zero frequencies of the spectrum are not perfectly aligned with the resulting discretization of the frequency domain, the convolution destroys much of the sparsity that existed in the continuous domain. Thus, considering the continuous Fourier transform permits to reduce the effect of this limitation and to recover the original sparsity in the light field. Here, the reconstruction is conducted in the angular domain, *i.e.* in (u, v)-coordinates, to reconstruct the missing samples (u, v).

The algorithm here operates in the intermediate domain $\hat{\mathscr{L}}_{w_x,w_y}(u, v)$ that describes spatial frequencies (w_x, w_y) as a function of viewpoint.

In a general context, a signal $x(t)$ of length N is k-sparse in the continuous Fourier domain if it can be represented as a combination of k non-discrete frequency coefficients:

$$x(t) = \frac{1}{N}\sum_{i=0}^{k} a_i exp(\frac{2\mathbf{j}\pi t w_i}{N}) \tag{3.1}$$

where $\{w_i\}_{i=0..k}$ are the non-discrete positions of frequencies, and $\{a_i\}_{i=0..k}$ are their corresponding coefficients (or values). Applying the previous equation to our light field data, the viewpoint

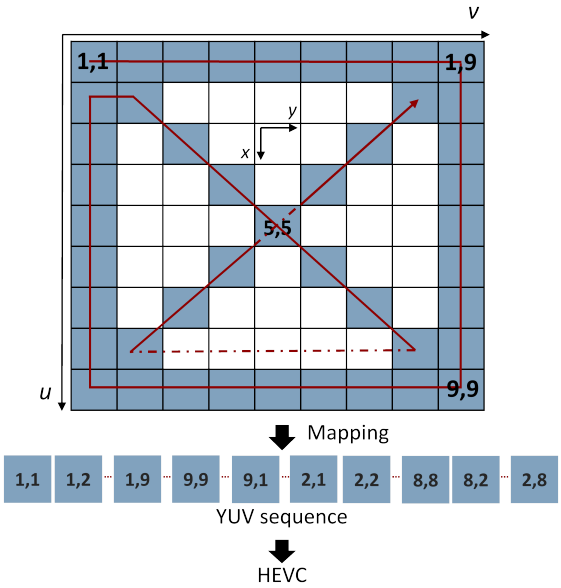


Figure 3.2: Selected sub-aperture images $\{I_{\mathbf{p}}\}_{\mathbf{p}\in P}$ sent as video frames following a specific scan order to HEVC encoder.

at position (u, v) of the signal L_{w_x,w_y} can be expressed as a linear combination of non-discrete angular frequency coefficients:

$$\mathscr{L}_{w_x,w_y}(u,v) = \sum_{(a_{u,v},w_u,w_v)} \frac{a_{u,v}}{N_a} exp(2\mathbf{j}\pi \frac{uw_u + vw_v}{N_a}), \tag{3.2}$$

where $\{w_u, w_v\}$ are the continuous frequency positions and $\{a_{u,v}\}$ are the corresponding coefficients. The objective is to recover the set $F = \{w_u, w_v, a_{u,v}\}$ of the sparse spectrum from the decoded views.

First, the 2D Fourier transform of each input viewpoint (u, v) from the decoded set $\{\hat{I}_{\mathbf{p}}\}_{\mathbf{p}\in P}$ is computed which gives the spatial frequencies (w_x, w_y) at the positions set P, consisting of the input 1D discrete lines. We will refer to this data as $\hat{\mathscr{L}}_{w_x,w_y}(u,v)_{|P}$, from which the 2D angular spectrum $\hat{\mathscr{L}}_{w_x,w_y}(w_u, w_v)$ for each spatial frequency (w_x, w_y) will be recovered.

The algorithm proceeds by first estimating integer frequency positions $\{w_u, w_v\}$ using a voting approach from the input sparse set. Then, the corresponding coefficients $\{a_{u,v}\}$ are estimated, and the frequency positions are refined to non-integer values using a two-step iterative approach. The method is detailed in Figure 3.3.

Initial Frequency Position Estimation The objective of this step is to initialize the set of positions $\{(w_u, w_v)\}$ of the non-zero frequency positions. Per the slicing theorem [17], the Fourier transform of a discrete line of a signal gives the projection of its spectrum onto that line.

Therefore, we calculate the Fourier transform in the angular domain (u, v) of each line segment of the input data $\mathscr{L}_{w_x,w_y}(u,v)_{|P}$. This yields the projections of the 2D spectrum $\mathscr{L}_{w_x,w_y}(w_u, w_v)$ on these lines.

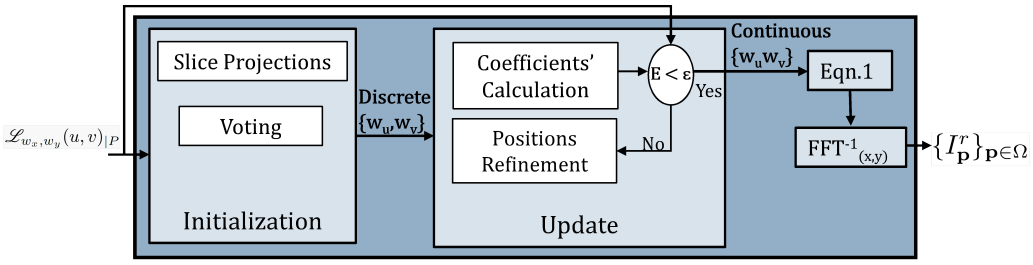


Figure 3.3: An overview of the sparse reconstruction scheme [22].

This corresponds to the step of *bucketization* [146] where the SFFT divides the frequency spectrum into buckets, and the value of each bucket represents the sum of the values of frequencies that map to it (see Figure.3.4). With the spectrum being sparse, many buckets will be empty, and thus the SFFT estimates the frequencies with large values in the non-empty buckets.

In our scheme, the *bucketization* is performed by a 1D DFT of each discrete line of the 1D selected viewpoint trajectory (the blue colored lines of viewpoints in Figure. 3.2). This yields the projection of the light field spectrum onto each line in the Fourier domain.

The frequency estimation then aims at identifying which frequencies created the energy in each bucket, and their corresponding values. For this purpose, a voting approach is used where each bucket votes for the frequencies that map to it. Due to the sparsity of the spectrum, only few frequencies receive a vote from every input projection, and will construct the initial estimation for angular frequency positions.

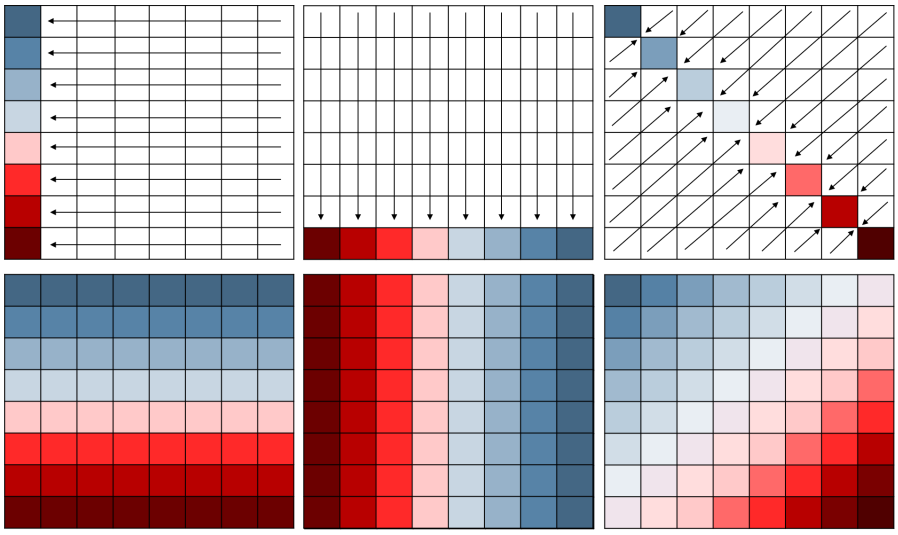


Figure 3.4: Frequency bucketization examples using discrete line Projections. Top: sampled discrete lines. Bottom: projection of the corresponding frequencies.

Coefficients Estimation and Frequency Refinement The algorithm proceeds afterwards by iteratively refining the coefficients and their frequency positions as follows:

- Given a set of positions $\{w_u, w_v\}$, the corresponding coefficients $\{a_{u,v}\}$ are recovered by solving the linear system of Eq. 3.2 over the set P of known (u, v) positions;
- Given the coefficients $\{a_{u,v}\}$, the frequency positions $\{w_u, w_v\}$ are refined to minimize the residual error $E(P)$, using a gradient descent algorithm based on finite differences

$$E(P) = \sum_{\mathbf{p} \in P} ||\hat{I}_{\mathbf{p}} - I^*_{\mathbf{p}}||^2, \tag{3.3}$$

where $\hat{I}_{\mathbf{p}}$ denotes the BL-decoded view at position $\mathbf{p}$ and $I^*_{\mathbf{p}}$ is the view at the same position, obtained by the sparse reconstruction algorithm.

The error threshold value is transmitted in the BL coder.

Based on the final frequency positions and corresponding coefficients, all the light field views are reconstructed using Eq. 3.2, followed by the inverse Fourier transform in (x, y) coordinates. This gives the reconstructed light field data $\{I^r_{\mathbf{p}}\}_{\mathbf{p} \in \Omega}$.

Quality Restoration

The quality of the reconstructed views $\{I^r_{\mathbf{q}}\}_{\mathbf{q} \in Q}$ may not be sufficient for some post-processing applications. We therefore consider these images, along with the images $\{\hat{I}_{\mathbf{p}}\}_{\mathbf{p} \in P}$, as inter-layer predictors of the original views in a scalable scheme which further encodes a prediction residue.

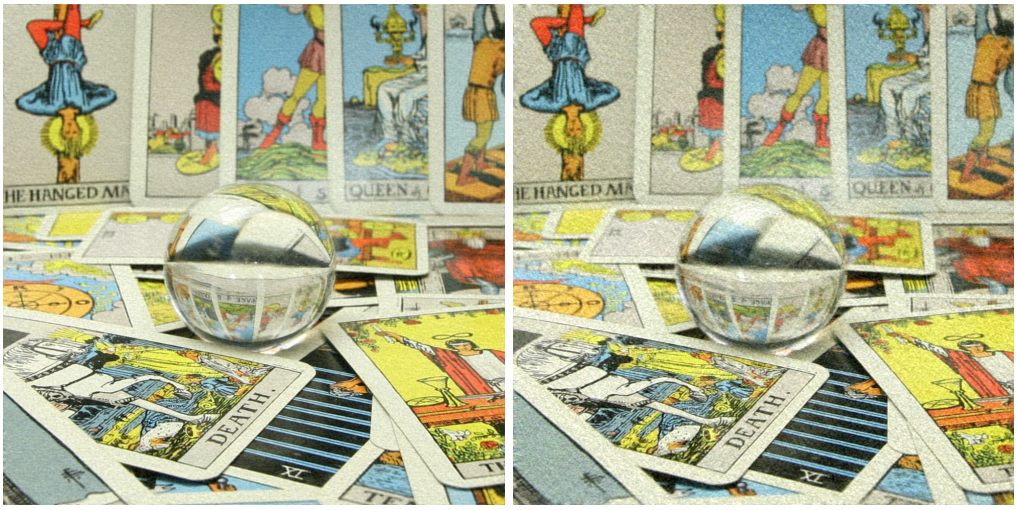
However, to ensure an efficient inter-layer prediction, the quality of the reconstructed images $\{I^r_{\mathbf{q}}\}_{\mathbf{q} \in Q}$ is first improved using a patch-based restoration method. Indeed, we have noticed that the resulting images from the SFFT-based method may suffer from some unstructured noise, due to the initialization step that includes a rough estimation of the frequency positions. Figure 3.5 shows an illustrative example of the reconstruction noise. Also, the farther a reconstructed sub-aperture image is from the input sample sub-aperture images, the lower its quality is.

We propose to restore the quality of these views using the BL-decoded views: we first assign the spatially closest BL-decoded image $\hat{I}_{\mathbf{p}}$ to each reconstructed image $I^r_{\mathbf{q}}$ as the reference image, and search with the PatchMatch algorithm [147] for the best matches between patches in the reference image and the image to be restored, and vice versa.

Once the matching process is over, the pixel values of the restored image are computed as a weighted average of overlapping patches, which minimizes the bidirectional similarity distance [148] applied between the reference image and the restored one.

To illustrate the pixel value calculation, let R_1,..., R_m denote all the patches in the reconstructed image $I^r_{\mathbf{q}}$ that contain a pixel s. D_1,...,D_m indicate the corresponding best matches in $\hat{I}_{\mathbf{p}}$, where d_1,...,d_m are the co-located pixels (see Figure 3.6).

Also, let $\hat{R}_1$,...,$\hat{R}_n$ denote all the patches in $I^r_{\mathbf{q}}$ that contain pixel s and serve as the best matches to $\hat{D}_1$,...,$\hat{D}_n$ in $\hat{I}_{\mathbf{p}}$, and $\hat{d}_1$,...,$\hat{d}_n$ be the co-located pixels in patches $\hat{D}_1$,...,$\hat{D}_n$ (see Figure 3.6).



view $(2, 3)$ (PSNR = 31.75dB) view $(5, 9)$ (PSNR = 28.81dB)

Figure 3.5: Examples of a reconstructed sub-aperture images from the *Crystal* light field using the Sparse Fourier Transform-based method in [22].

Hence, the value of s in the final restored image $I_{\mathbf{q}}^{r*}$ is expressed as

$$I_{\mathbf{q}}^{r*}(s) = \frac{\sum_{i=1}^{m} \hat{I}_{\mathbf{p}}(d_i) + \sum_{j=1}^{n} \hat{I}_{\mathbf{p}}(\hat{d}_j)}{m+n}. \tag{3.4}$$

Figure 3.7 shows an example of restoration result of a reconstructed view from the *Crystal* dataset: the zoom on the extracted areas demonstrates that our method provides a good solution to remove the noise from SFFT reconstruction results.

An average PSNR gain of 2.8dB is obtained for the dataset *Crystal*. One can also note that the restoration does not alter the structures in the light field images, and does not require more input details, but only the BL-decoded sub-aperture images.

3.2.2 Enhancement Layer Coding

The restored views $\{I_{\mathbf{q}}^{r*}\}_{\mathbf{q}\in Q}$ and the BL-decoded $\{\hat{I}_{\mathbf{p}}\}_{\mathbf{p}\in P}$ are loaded into the inter-layer reference picture list in a SNR-scalable coder in SHVC, for the prediction of the original light field.

Both inter-layer and *intra/inter* predictions are performed during encoding, and the best coding mode is chosen (with Rate Distortion Optimization) for each block. Residues from prediction are then quantized, transformed and encoded, delivering the enhancement layer (EL) bitstream. A final full light field data is delivered by the decoder, using both BL and EL bitstreams.

The proposed scheme is scalable and has two layers. The first layer is the sparse sampled set of sub-aperture images, coded and reconstructed using SFFT. The reconstruction quality depends on the sampling factor. The second layer is the residues from the prediction using the reconstructed light field.

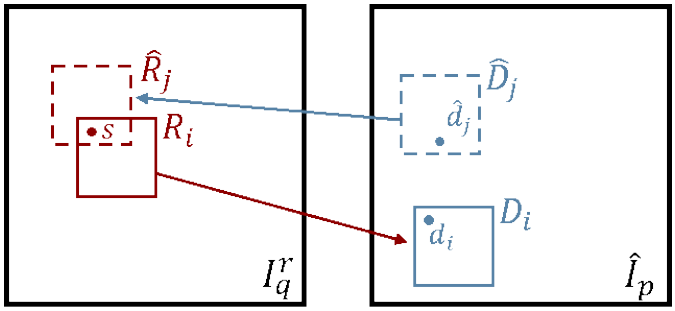


Figure 3.6: Pixel value restoration based on bidirectional matching using *PatchMatch* algorithm [147].

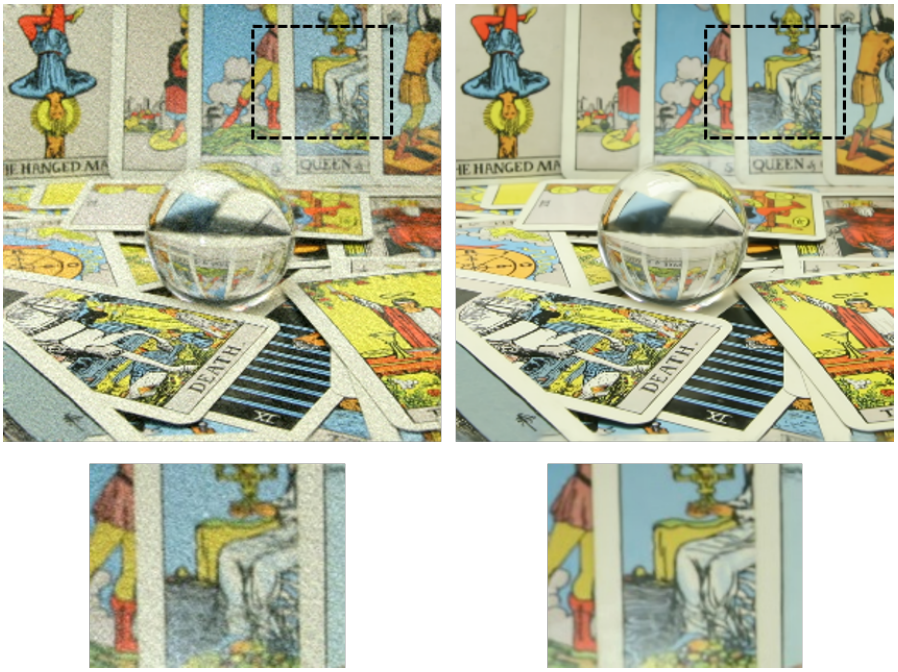


Figure 3.7: An example of a reconstructed image $I_{\mathbf{q}}^r$ (left) and its corresponding restored image $I_{\mathbf{q}}^{r*}$ (right) from *Crystal* dataset. Note that the presence of heavy noise in the reconstructed image does not allow its use as an inter-layer predictor to enhance the compression efficiency of the EL.

This scalability property is beneficial if the resource in the network is limited, since the set of images in the base layer is smaller than the full original light field, and is sufficient to produce and 4D light field if the sampling rate is appropriate. Also, the base layer can be set to a higher priority transmission in the case of differentiated network, where it is always possible to reconstruct the full light field content from the base layer.

3.3 Experimental Setup and Rate-distortion Results

In our experiments, the HEVC Test Model (HM 16.9) software is used for the coding of sub-aperture images in the base layer (BL), with a GOP size of 8 in a hierarchical structure. Only the upper-left view of the light field matrix is intra-encoded. The QP values are set to 22, 27, 32 and 37.
The matrix of views is scanned row-by-row from left to right and from right to left to form a video sequence, which is converted to YUV (4:2:0). This sequence is then encoded using the SHVC Test Model (SHM 12.1) with the same parameters as above, yielding the enhancement layer (EL) bitstream. We compare our compression scheme with a HEVC single layer (SL)

coding of the original matrix of views converted into a YUV sequence, in the same way as the EL sequence. Various light field datasets are used in the evaluation tests. First, synthetic light fields [1] *Buddha*, *Still Life*, *Butterfly* and *Horses* are 9x9 sub-aperture images of a resolutions of 768x768 or 1024x576 RGB pixels each. Real multi-view light fields with large disparity are also used from the *Stanford* dataset [2]: each light field is composed of 17x17 sub-aperture images of large spatial resolution of the order of more than 1 MP. Besides, plenoptic images from the JPEG Pleno Dataset [149] are tested with our compression scheme: they are captured with the Lytro camera, and compose of 13x13 sub-aperture images of 625x434 spatial resolution. Following the selection pattern for input images of the base layer, only 45 views are retained for the synthetic dataset, 93 views out of 289 for the multi-view captured light fields, and finally 69 out of 169 for the plenoptic contents.

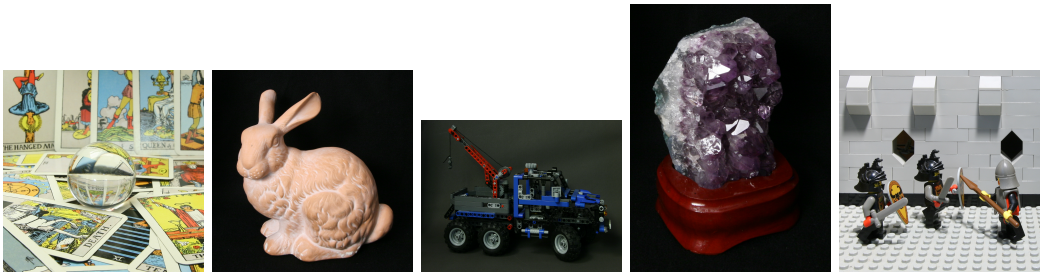
The objective quality is assessed on the YUV components with PSNR averaged on all the viewpoints for each light field, and the bit-rate is calculated from the coded bitstream for all YUV components. Average rate-distortion curves are plotted in Figure 3.11 for each type of light field dataset.

The results in Figure 3.11 show that significant gains are obtained by the proposed compression scheme, compared to direct coding if the light field images in a HEVC single layer, especially in high bit-rate. Furthermore, rate-distortion performances are computed with the *Bjontegaard* metrics [150]. The results obtained for different light fields are presented in Table 3.1. Our scalable compression scheme can improve the coding efficiency by up to 24% bit-rate saving compared to HEVC single layer coding. Important gains, of an of average 17.9% of bit-rate reduction, are achieved in the case of multi-view light fields (from the Stanford dataset) that exhibit large baselines, which validates the interest of the proposed method for the compression of high-resolution light fields.

[1] http://lightfieldgroup.iwr.uni-heidelberg.de
[2] http://lightfield.stanford.edu/lfs.html

Buddha StillLife Butterfly Horses

Figure 3.8: Tested synthetic light fields from HCI dataset.



Crystal Bunny Lego Truck Amethyst Lego Knights

Figure 3.9: Tested light fields from the Stanford dataset.

Bikes Friends 1 StonePillars Vespa

Figure 3.10: Tested plenoptic light fields from the JPEG Pleno dataset.

3.4 Impact on a Light Field Application: Extended Field of Focus

In this section, we aim at analyzing the impact of compression on a post-capture image rendering application: generation of the extended focus image. While the focus stack images correspond to different focusing plans in the scene, the extended focus image is generated by selecting for each pixel the focus stack image where it is in-focus.

A refocused image after compression contains both the blur due to compression and the refocusing blur. In out-of-focus areas, the quantization blur is mixed with the geometry blur, which makes it difficult to evaluate the compression impact in these images. The compression

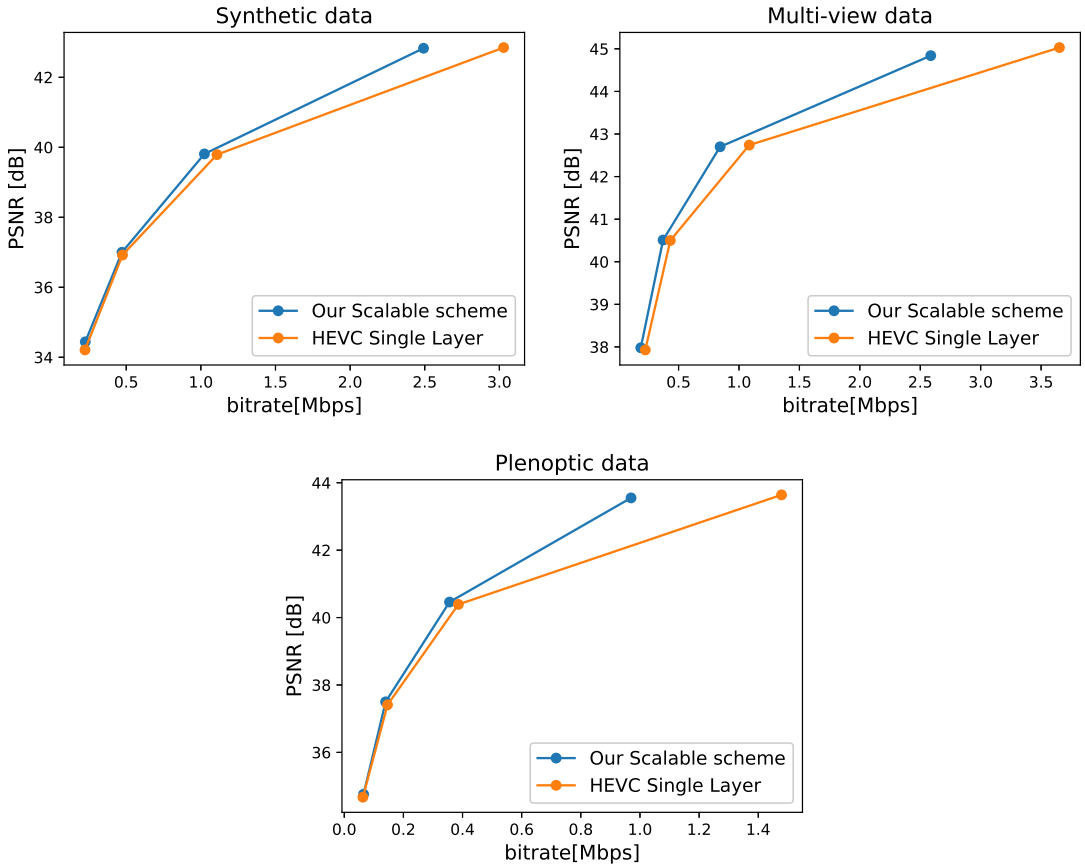


Figure 3.11: Rate-distortion performance of our compression scheme compared to HEVC single layer coding for different light field contents.

blur is in contrast visible in the extended focus images. Thus, we propose to measure the quantization blur in the extended focus images resulting from the compression scheme. The evaluation is made only on HCI datasets for which ground truth depth maps are available. We avoid in this case any quality alteration that may come from depth estimation errors. In this case, only the compression blur propagates from in-focus areas of the focus stack to the extended focus image.

We first compute the focus stack images $\{\mathfrak{F}_{\alpha_i}\}$ by performing a digital refocusing of the central view [151], at different depth values α_i extracted from the depth map $D(x, y)$. The depth of field can then be extended. The correspondent all-in-focus image $E(x, y)$ is obtained by choosing, for a pixel at position (x, y) of depth α_i, the pixel of the image from the focus stack which is refocused at this depth:

$$E(x, y) = \mathfrak{F}_{\alpha_i}(x, y) \tag{3.5}$$

Figure 3.12 illustrates the generation of this all-in-focus image.

Table 3.1: Bit-rate savings and PSNR gains compared to HEVC single layer coding.

		BD-rate(%)	BD-PSNR(dB)
HCI dataset	Buddha	-11.83	0.50
	StillLife	-8.17	0.30
	Butterfly	-4.70	0.21
	Horses	-4.90	0.16
Stanford dataset	Crystal	-24.24	0.67
	Bunny	-22.23	0.58
	Lego Truck	-18.00	0.60
	Amethyst	-17.55	0.52
	Lego Knights	-7.50	0.30
Plenoptic dataset [149]	Bikes	-8.42	0.30
	Friends 1	-5.64	0.19
	StonePillarsOutside	-3.62	0.11
	Vespa	-8.90	0.28

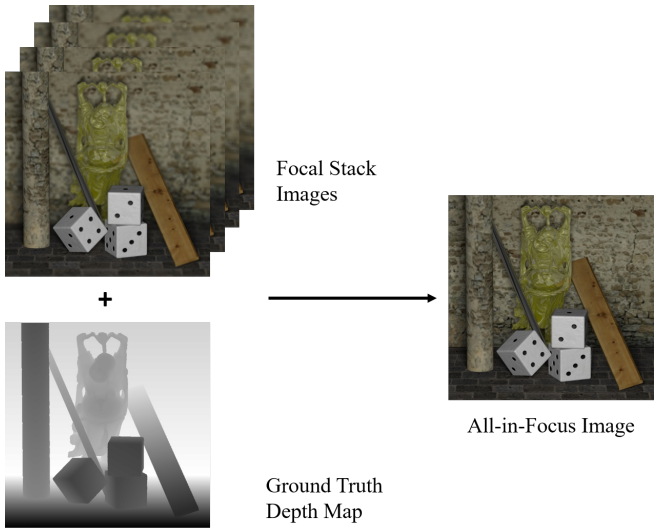


Figure 3.12: Generating the all-in-focus image using the focus stack and the ground truth depth map.

As the compression blur alters the perceived quality of images, traditional metrics such as PSNR fail to accurately measure the impact resulting from compression on refocused images, as it has been shown in [69]. We use here a perceptual edge-based blur metric, from [152]. We compare the quality of the all-in-focus image resulting from the outputs of each compression scheme to the one of the original light field. A higher blur measure denotes a degraded quality. Figure 3.13 shows the results of blur measurement using Marziliano metric [152] as a function of the compression bit-rate. It demonstrates that the proposed scheme outperforms HEVC single layer

coding and introduces less distortion in decoded images. High compression ratios are achieved, without altering the visual quality of the all-in-focus image.

The plotted results align with our visual perception. For the few highest QP (most compressed data), a significantly degraded quality of the extended focus, the focus stack and the reconstructed light field is observed. However, increasing progressively the bitrate will lead to an almost stable quality in terms of blur.

Besides, it can be noted that the output of the reconstruction based on BL-decoded views (illustrated in green) provides an equivalent quality of the all-in-focus image as the HEVC single layer coding, but at a significantly lower compression cost.

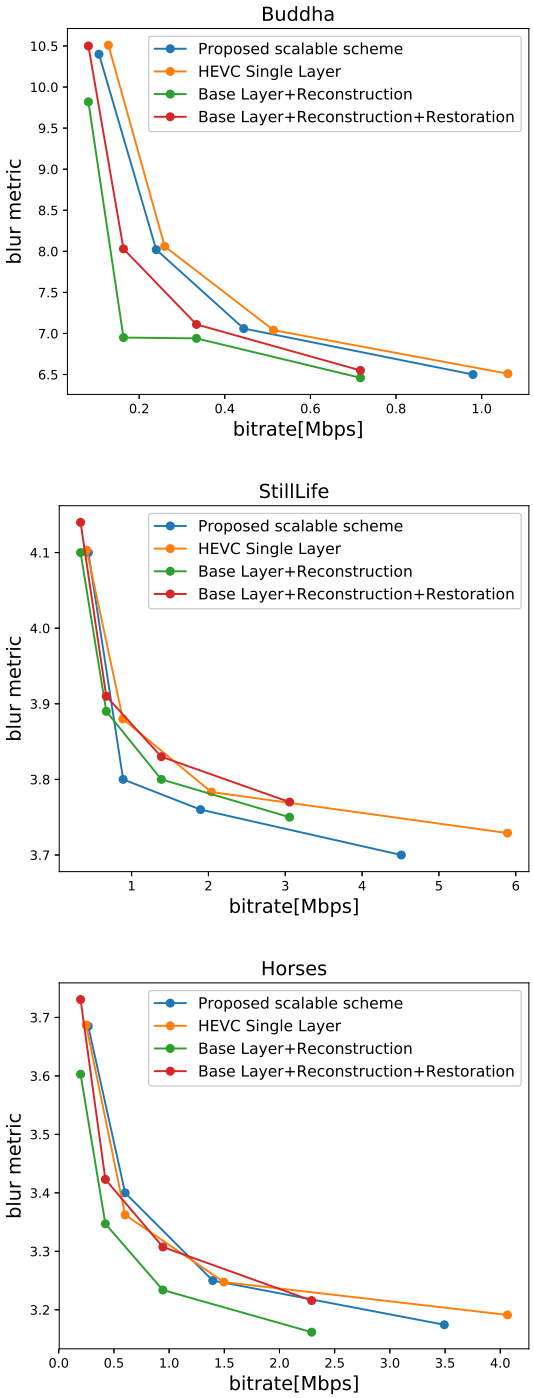


Figure 3.13: Amount of blur in the all-in-focus image resulting from different output data: a higher blur measure indicates a lower quality of the all-in-focus image.

3.5 Conclusion

We introduced a scalable coding scheme for light field data. A sparse viewpoint subset is selected and encoded as a base layer with HEVC. A full light field is reconstructed from the decoded images using a sparse recovery method in the Fourier domain. These reconstructed data are enhanced using the decoded views, and then used as prediction reference for inter-layer coding of the entire original light field.

The proposed scheme is scalable with two layers, so that the data used for rendering can either consist of the plain reconstruction from the sparse view samples, or its refined version with the enhancement layer. Experimental results demonstrate that this scalable scheme outperforms HEVC single layer encoding for the tested light field datasets, both synthetic and real. The analysis of all-in-focus images also shows that our method does not induce visual artifacts, even for data reconstructed from the base layer, which presents an advantageous outcome for further post-capture light field applications.

While several works have focused on efficient compression schemes for the transmission of light field, the acquisition of large densely sampled light fields (especially light field videos) still presents a challenging problem as well, due to the limited acquisition and storage facilities. We propose to tackle in details this issue in the next chapter.

Chapter 4

A Sparsity-based Reconstruction of Sub-sampled Light Fields

4.1 Introduction

While the industry calls for increasing resolutions and frame-rates, the task of acquiring high-quality 4D content remains challenging due to the complexity and size of optics, photo-sensors, and ultimately, because of the bottleneck of data storage. Indeed, capturing light field images and videos requires sophisticated systems and engineering prodigality to operate the incoming data stream. When distributed storage is not an option, *e.g.* for real-time pre-visualization purpose, light field video acquisition setups have no choice but sacrificing the resolution in one or several dimension: spatial, angular, or temporal.

Nevertheless, light field photography keeps on gaining popularity as the need for immersive acquisitions increases in the Entertainment industry. Yet, light field data often exhibit insufficient sampling in one dimension or another, either because of the physical limitations of the acquisition system itself, or because of the performances of the storage facilities.

To treat the aforementioned problem, we present in this chapter a novel automatic reconstruction method of light fields from sparse data samples, based on Orthogonal Frequency Selection (OFS) in the Fourier domain. A sub-sampled version of the captured data is stored, then used to restore the full-resolution light field. The method is independent to the acquisition system, and does not require any prior knowledge of the scene geometry or any pre-processing technique to estimate the depth information.

4.2 Overview of the Reconstruction Method

4.2.1 Problem Statement

Let $\mathcal{L}_s(x, y, u, v)$ be a 4D randomly-sampled light field. We aim at estimating the full 4D signal $\mathcal{L}_r(x, y, u, v)$ that best fits the available samples of $\mathcal{L}_s(x, y, u, v)$ with a sufficient quality level. The desired signal $\mathcal{L}(x, y, u, v)$ is regarded as sparse in the Fourier domain, and thus can be described as

$$\mathcal{L} = \Psi\alpha, \tag{4.1}$$

with Ψ being the matrix containing all the possible Fourier basis functions, forming a complete Fourier basis, and α being the sparse vector of expansion coefficients.

The available signal samples result from the sub-sampling of the unavailable signal $\mathcal{L}$ at non-regularly spaced positions. The generation process of these samples can be described as

$$\mathcal{L}_s = \Phi\mathcal{L} = \Phi\Psi\alpha, \tag{4.2}$$

where Φ represents the sub-sampling matrix containing 0 and 1 values to determine the available signal after sub-sampling. We aim to generate the sparse model $\mathcal{L}_r$ which can be defined as

$$\mathcal{L}_r = \Psi\hat{\alpha}. \tag{4.3}$$

The coefficients in $\hat{\alpha}$ correspond to the contributions of the frequencies that our reconstruction method selects iteratively based on the available samples in $\mathcal{L}_s$. Examples of sub-sampling at different rates are shown in Figure 4.2. Each viewpoint(or angular sample) of the light field is spatially sub-sampled following the chosen sampling rate, independently of the other viewpoints.

4.2.2 Sparse Model for Light Field Reconstruction

We conduct the reconstruction per 4D hyper-block of the light field, and an approximation model $g[k, l, m, n]$ is generated for each hyper-block signal $f[k, l, m, n]$; (k, l) and (m, n) correspond respectively to the spatial and angular coordinates of the signal within a 4D hyper-block.

If a hyper-block is to be reconstructed, it is always regarded as the core of a surrounding area in the spatial and angular dimensions. An example is illustrated in Figure 4.1. This neighboring area is composed of a stripe of samples of a defined border width in spatial and angular directions. Together with its local surroundings, the hyper-block forms a 4D domain Ω spanning over $M \times N$ views with co-located $K \times L$ patches.

Let A, B and C respectively denote the local subsets of the known, unknown, and reconstructed samples: $\Omega = A \cup B \cup C$. As neighboring hyper-blocks may already have been processed before, the corresponding reconstructed samples are contained in the area C to be used in the reconstruction of the currently considered hyper-block.

For convenience, we summarize the notations used throughout this chapter in Table 4.1.

Table 4.1: Notations

$\mathbf{j}$	The imaginary unit: $\mathbf{j}^2 = -1$
$z^* = \Re(z) - \mathbf{j}.\Im(z)$	Complex conjugate of z
i	Iteration index
$\Omega = [\![1; K]\!] \times [\![1; L]\!] \times [\![1; M]\!] \times [\![1; N]\!]$	Local light field domain
$P = \lvert\Omega\rvert = K.L.M.N$	Number of samples in Ω
$A \subset \Omega$	Subset of known samples
$B \subset \Omega$	Subset of unknown samples
$C \subset \Omega$	Subset of reconstructed samples
$\mathbf{p} = (k, l, m, n) \in \Omega$	A pixel's position within Ω
$f : \Omega \to \mathbb{R}$	Local light field
$g^{(i)} : \Omega \to \mathbb{R}$	Approximation model at iteration i
$w : \Omega \to \mathbb{R}$	Weighting function
$r^{(i)} : \Omega \to \mathbb{R}$	Weighted residue at iteration i
$\boldsymbol{\vartheta} = (\mu, \nu, \zeta, \xi)$	A frequency in the 4D spectrum
$\varphi_{\boldsymbol{\vartheta}} : \Omega \to \mathbb{R}$	A 4D Fourier basis function
$\Theta^{(i)} = \{\boldsymbol{\vartheta}_1, \ldots, \boldsymbol{\vartheta}_i\}$	Frequency subset at iteration i
$X_{\boldsymbol{\vartheta}} = \sum_{\mathbf{p}} x[\mathbf{p}]\varphi^*_{\boldsymbol{\vartheta}}[\mathbf{p}] \in \mathbb{C}$	Capitals denote Fourier transforms

Besides, a weighting function w is introduced to discriminate known (*i.e.* original) and reconstructed samples from unknown samples in Ω:

$$w[\mathbf{p}] = w[k, l, m, n] = \begin{cases} \rho_s^{\sqrt{\overline{k}^2 + \overline{l}^2}} \rho_a^{\sqrt{\overline{m}^2 + \overline{n}^2}} & \text{for } \mathbf{p} \in A \\ 0 & \text{for } \mathbf{p} \in B \\ \sigma.\rho_s^{\sqrt{\overline{k}^2 + \overline{l}^2}} \rho_a^{\sqrt{\overline{m}^2 + \overline{n}^2}} & \text{for } \mathbf{p} \in C \end{cases} \tag{4.4}$$

where $\overline{k} = k - \frac{K+1}{2}$, $\overline{l} = l - \frac{L+1}{2}$, $\overline{m} = m - \frac{M+1}{2}$, $\overline{n} = n - \frac{N+1}{2}$, and $0 < \rho_s, \rho_a, \sigma < 1$. The factors ρ_s and ρ_a respectively determine the weighting decay in the spatial dimension (within a view) and in the angular (cross-view) dimension. The weighting function w is used to control the influence each sample has on the reconstruction process depending on its position in area Ω. In this sense, the weight of a sample decreases as its distance to the hyper-block center increases. The unavailable samples are masked with w since we do not use their contributions in the approximation process. Finally, the parameter σ is used to differentiate the contribution of reconstructed data from the available data.

Let f be the local light field defined on Ω. Taking advantage of the sparsity of light field contents in the Fourier domain, the algorithm generates a sparse approximation model g of the signal f:

$$g[\mathbf{p}] = \sum_{\boldsymbol{\vartheta} \in \Theta} c_{\boldsymbol{\vartheta}}.\varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] \tag{4.5}$$

as a weighted combination of Fourier basis functions defined as:

$$\varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] = \varphi_{\mu\nu\zeta\xi}[k, l, m, n] = e^{2\pi\mathbf{j}\left(\frac{k\mu}{\mathbf{K}} + \frac{l\nu}{\mathbf{L}} + \frac{m\zeta}{\mathbf{M}} + \frac{n\xi}{\mathbf{N}}\right)} \tag{4.6}$$

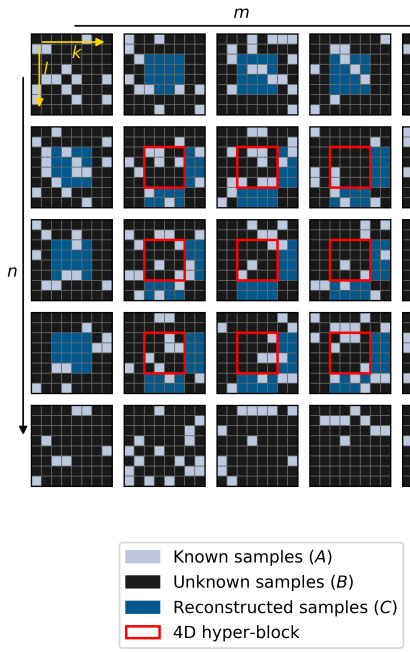


Figure 4.1: 4D hyper-block (outlined in red) and its local spatio-angular neighborhood ($K = L = 8$ and $M = N = 5$).

4.2.3 Iterative Reconstruction by Orthogonal Frequency Selection

The main idea of our reconstruction approach consists in selecting the best Fourier basis functions that would represent the signal within the currently considered hyper-block, using only the available data we dispose of. The selection is carried out iteratively in a way that minimizes the residual error at each iteration, inspired from [1].

This is similar to the Matching Pursuit algorithm [2] in the sense that it sparsely approximates a signal by finding the best matching projections of the data onto the span of a Fourier basis. The different reconstruction steps of our method are outlined in the representative scheme in Figure 4.3. We will explain below each step in details.

Since the reconstruction of a hyper-block greatly depends on the information available in its neighborhood, the choice of reconstruction order can be adapted to the data state at each step. For optimizing the order in which the individual hyper-blocks are processed, we need to take into account some properties of the approximation algorithm we use. First of all, extrapolation quality generally increases with an increasing number of known samples. Due to the reuse of already extrapolated samples, this also means that a hyper-block should not be processed until as many as possible of its neighboring hyper-blocks are available.

Since the blocks are reconstructed one after the other, using known or already reconstructed neighboring samples, the updated local density of the neighboring area is an important parameter

to take into account to choose the next block to restore. The original order is based on the initial local density of each individual block, which does not take advantage of the higher density of the already reconstructed blocks around.

In practice, we assign to each hyper-block a surrounding local density level that will be updated (increased) each time one of its neighbors is reconstructed, increasing its label in the global reconstruction order of the light field. The surrounding density $sd(b)$ of a hyper-block b taking into account the densities d of its neighbors $\{b_x \in Nb(b)\}$ is then described as:

$$sd(b) = d(b) + \sum_{b_x \in Nb(b)} d(b_x) \tag{4.7}$$

In the reconstruction algorithm, to choose the following block to reconstruct, we rely on the updated local densities of all non-reconstructed blocks. This ensures a better local coherence, and higher approximation quality, since we ensure a higher density of known samples at each hyper-block reconstruction. As such, the extrapolation result from one hyper-block will be used to improve the model generation of its sampled neighbors.

Figure 4.2: Examples of sampled images of the light field dataset *Cars* at different sampling rates. Top to bottom: original image, 40%-sampling, 20%-sampling and 10%-sampling.

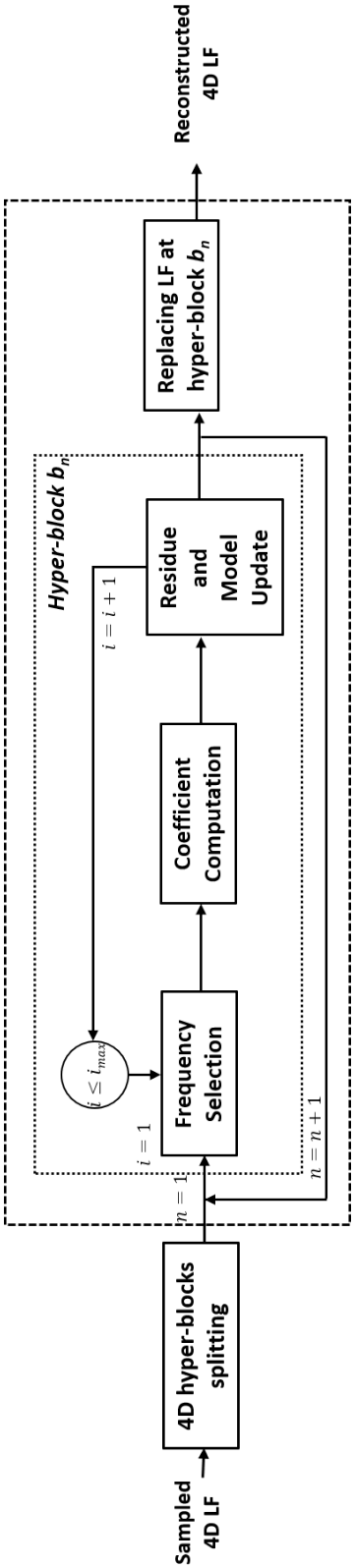


Figure 4.3: Representative scheme of our proposed compressive light field reconstruction method.

Frequency Selection

We aim at iteratively selecting the best basis functions to approximate the original signal, and thus maximally reducing the residual error to the available samples at each iteration. Let $r^{(i)}$ be the weighted residue of the approximation model with respect to the signal at iteration i:

$$r^{(i)} = (f - g^{(i)}).w \tag{4.8}$$

The selected basis function $\boldsymbol{\vartheta}_i$ at iteration i is the one on which the projection of the residue $r^{(i-1)}$ is maximal:

$$\begin{aligned} \boldsymbol{\vartheta}_i &= \underset{\boldsymbol{\vartheta}}{\operatorname{argmax}} \left| \langle \varphi_{\boldsymbol{\vartheta}}, r^{(i-1)} \rangle \right| \\ &= \underset{\boldsymbol{\vartheta}}{\operatorname{argmax}} \left| \sum_{\mathbf{p} \in \Omega} \left(r^{(i-1)}[\mathbf{p}].\varphi^*_{\boldsymbol{\vartheta}}[\mathbf{p}] \right) \right| \end{aligned} \tag{4.9}$$

Update with No Orthogonality Constraint

The expansion coefficient $c^{(i)}_{\boldsymbol{\vartheta}_i}$ shall minimize the weighted residual energy $E^{(i)}_w$, as proposed in [1]:

$$E^{(i)}_w = \sum_{\mathbf{p} \in \Omega} \left| f[\mathbf{p}] - g^{(i-1)}[\mathbf{p}] - c^{(i)}_{\boldsymbol{\vartheta}_i}.\varphi_{\boldsymbol{\vartheta}_i}[\mathbf{p}] \right|^2 .w[\mathbf{p}] \tag{4.10}$$

Looking for zeros of the derivative of $E^{(i)}_w$ with respect to $c^{(i)}_{\boldsymbol{\vartheta}_i}$ yields

$$c^{(i)}_{\boldsymbol{\vartheta}_i}. \sum_{\mathbf{p} \in \Omega} \left(w.\varphi_{\boldsymbol{\vartheta}_i}.\varphi^*_{\boldsymbol{\vartheta}_i} \right)[\mathbf{p}] = \sum_{\mathbf{p} \in \Omega} \left(r^{(i-1)}.\varphi^*_{\boldsymbol{\vartheta}_i} \right)[\mathbf{p}] \tag{4.11}$$

And therefore:

$$c^{(i)}_{\boldsymbol{\vartheta}_i} = \frac{\sum_{\mathbf{p} \in \Omega} r^{(i-1)}[\mathbf{p}].\varphi^*_{\boldsymbol{\vartheta}_i}[\mathbf{p}]}{\sum_{\mathbf{p} \in \Omega} w[\mathbf{p}]} \tag{4.12}$$

At that point, a straightforward update step of both the approximation model and the weighted residue would consist in:

$$\begin{cases} g^{(i)} = g^{(i-1)} + c^{(i)}_{\boldsymbol{\vartheta}_i}.\varphi_{\boldsymbol{\vartheta}_i} \\ r^{(i)} = r^{(i-1)} - c^{(i)}_{\boldsymbol{\vartheta}_i}.\varphi_{\boldsymbol{\vartheta}_i}.w \end{cases} \tag{4.13}$$

Yet, while the Fourier basis functions are obviously orthogonal on Ω, a weighted basis function is *a priori* not orthogonal to another basis function:

$$\forall\ \boldsymbol{\vartheta} \neq \boldsymbol{\vartheta}'\ ,\ \langle w.\varphi_{\boldsymbol{\vartheta}}, \varphi_{\boldsymbol{\vartheta}'} \rangle \neq 0 \tag{4.14}$$

Therefore, if the update step (4.13) cancels the energy of $r^{(i-1)}$ along $\boldsymbol{\vartheta}_i$ in $r^{(i)}$, it also alters its spectrum for every $\boldsymbol{\vartheta}_{j<i}$:

$$\begin{cases} \langle r^{(i)}, \varphi_{\boldsymbol{\vartheta}_i} \rangle = 0 \\ \forall\ j < i\ :\ \boldsymbol{\vartheta}_j \neq \boldsymbol{\vartheta}_i\ ,\ \langle r^{(i)}, \varphi_{\boldsymbol{\vartheta}_j} \rangle \neq 0 \end{cases} \tag{4.15}$$

The weighted residue is not orthogonal to the subspace spanned by the already selected functions. As a consequence, a basis function can possibly be selected repeatedly. To palliate this orthogonality deficiency, Seiler *et al.* [1] rely on a compensation factor between 0 and 1 in the computation of $c_{\boldsymbol{\vartheta}_i}$ to re-sample 2D sub-sampled images.

Still, their approach requires many iterations to reach a fair approximation quality, for example, 100 iterations for 4×4 blocks and 32×32 neighborhood [1]. Moreover, the value of the compensation factor is derived from a trained parameter. As such, a suitable dataset needs to be selected for the training, which may limit the method performances when applied on other types of data.

For the orthogonality to be properly enforced, the minimization criterion should take into account the whole approximation model spectrum at each iteration. This idea has been discussed in the context of image coding in [136] under the name of *best approximation*, but has never been analyzed, nor used before to sparsely reconstruct randomly sampled light fields.

Orthogonality and Hermitian Symmetry

To better estimate the contribution of each newly selected function, we introduce a new criterion to approximate the residual error, taking into account all the basis functions selected so far. By incorporating all the selected basis functions in the update of the approximated residue, a more accurate reconstruction can be obtained with a reduced number of iterations:

$$\begin{cases} g^{(i)} = g^{(i-1)} + \sum_{\boldsymbol{\vartheta} \in \Theta^{(i)}} \Delta c_{\boldsymbol{\vartheta}}^{(i)} . \varphi_{\boldsymbol{\vartheta}} \\ r^{(i)} = r^{(i-1)} - \sum_{\boldsymbol{\vartheta} \in \Theta^{(i)}} \Delta c_{\boldsymbol{\vartheta}}^{(i)} . \varphi_{\boldsymbol{\vartheta}} . w \end{cases} \tag{4.16}$$

The basis function selection is conducted in the same way as explained before. However, the new minimization criterion with respect to each $\Delta c_{\boldsymbol{\vartheta}}$ changes the equation (4.10) to:

$$E_w^{(i)} = \sum_{\mathbf{p} \in \Omega} \left| f[\mathbf{p}] - g^{(i-1)}[\mathbf{p}] - \sum_{\boldsymbol{\vartheta} \in \Theta^{(i)}} \Delta c_{\boldsymbol{\vartheta}}^{(i)} . \varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] \right|^2 . w[\mathbf{p}] \tag{4.17}$$

Also, as real-valued light field signals are Hermitian, their Fourier spectra show conjugate complex symmetries. So, to ensure that the approximation g yields a real-valued signal, we modify (4.5) to:

$$g[\mathbf{p}] = \frac{1}{2} \sum_{\boldsymbol{\vartheta} \in \Theta} \left(c_{\boldsymbol{\vartheta}} . \varphi_{\boldsymbol{\vartheta}}[\mathbf{p}] + c_{\boldsymbol{\vartheta}}^* . \varphi_{\boldsymbol{\vartheta}}^*[\mathbf{p}] \right) \tag{4.18}$$

This means adding to the model the conjugate complex of a frequency each time it is selected:

$$E_w^{(i)} = \sum_{\mathbf{p} \in \Omega} \left| \left(f - g^{(i-1)} - \frac{1}{2} \sum_{\boldsymbol{\vartheta} \in \Theta^{(i)}} \left(\Delta c_{\boldsymbol{\vartheta}}^{(i)} . \varphi_{\boldsymbol{\vartheta}} + \Delta c_{\boldsymbol{\vartheta}}^{(i)*} . \varphi_{\boldsymbol{\vartheta}}^* \right) \right)[\mathbf{p}] \right|^2 . w[\mathbf{p}] \tag{4.19}$$

The new minimization criterion with respect to each $\Delta c_{\boldsymbol{\vartheta}}^{(i)}$ yields a system of equations whose solution ensures that the residue is orthogonal to each basis function selected so far:

$$\begin{aligned}&\sum_{\mathbf{p}\in\Omega}\Big(w[\mathbf{p}].\varphi_{\boldsymbol{\vartheta}}[\mathbf{p}]\sum_{\boldsymbol{\vartheta'}\in\Theta^{(i)}}\frac{1}{2}\big[\Delta c_{\boldsymbol{\vartheta'}}^{(i)}.\varphi_{\boldsymbol{\vartheta'}}[\mathbf{p}]+\Delta c_{\boldsymbol{\vartheta'}}^{(i)*}.\varphi_{\boldsymbol{\vartheta'}}^{*}[\mathbf{p}]\big]\Big)\\&=\sum_{\mathbf{p}\in\Omega}r^{(i-1)}[\mathbf{p}].\varphi_{\boldsymbol{\vartheta}}[\mathbf{p}],\quad\forall\boldsymbol{\vartheta}\in\Theta^{(i)}\end{aligned}\tag{4.20}$$

Approximation Model and Residue Update

Once a new basis function is selected, an update step introduces the contribution of the selected function basis to the approximation model, and updates to the contributions of the already selected functions:

$$g^{(i)}=g^{(i-1)}+\frac{1}{2}\sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}}\big(\Delta c_{\boldsymbol{\vartheta}}^{(i)}.\varphi_{\boldsymbol{\vartheta}}+\Delta c_{\boldsymbol{\vartheta}}^{(i)*}.\varphi_{\boldsymbol{\vartheta}}^{*}\big)\tag{4.21}$$

As for the residue, the contribution of the selected basis function $\varphi_{\boldsymbol{\vartheta}_i}$ is removed, as well as all the weighted updates of the previously selected functions:

$$r^{(i)}=r^{(i-1)}-\frac{1}{2}\sum_{\boldsymbol{\vartheta}\in\Theta^{(i)}}\big(\Delta c_{\boldsymbol{\vartheta}}^{(i)}.\varphi_{\boldsymbol{\vartheta}}+\Delta c_{\boldsymbol{\vartheta}}^{(i)*}.\varphi_{\boldsymbol{\vartheta}}^{*}\big).w\tag{4.22}$$

The algorithm then proceeds to the next iteration where a new basis function is selected, and so on, until a predefined number of iterations is reached.

4.2.4 Analytical Solution in the Fourier Domain

So far, the reconstruction method has been explained in the spatial domain. Using DFT functions as basis functions allows us to express all the equations in the frequency domain. With the Fourier transform, the evaluation of the sum in (4.20) can be less computationally expensive than in the spatial domain, and an efficient implementation of the reconstruction algorithm can be obtained. Only one local Discrete Fourier Transform (DFT) at the beginning and one inverse transform at the end are necessary, all intermediate steps being expressed in the Fourier domain.

The frequency selection step (4.8) can be expressed as follows in the Fourier domain:

$$\boldsymbol{\vartheta}_i=\underset{\boldsymbol{\vartheta}}{\operatorname{argmax}}\left|R_{\boldsymbol{\vartheta}}^{(i-1)}\right|\tag{4.23}$$

where R denotes the Fourier transform of the weighted residue r.

As for the expansion coefficients computation, the system of equations (4.20) becomes, for every $\boldsymbol{\vartheta}\in\Theta^{(i)}$:

$$\sum_{\boldsymbol{\vartheta'}\in\Theta^{(i)}}\frac{1}{2}\big(\Delta c_{\boldsymbol{\vartheta'}}.W_{\boldsymbol{\vartheta'}+\boldsymbol{\vartheta}}^{*}+\Delta c_{\boldsymbol{\vartheta'}}^{*}.W_{\boldsymbol{\vartheta'}-\boldsymbol{\vartheta}}\big)=R_{\boldsymbol{\vartheta}}^{(i-1)^*}\tag{4.24}$$

where W denotes the Fourier transform of the weighted function w.

Now, the objective is to estimate the updates of all the expansion coefficients $\{\Delta c^{(i)}_{\boldsymbol{\vartheta}_j}\}_{j=1..i}$, *i.e.* to solve the following equation:

$$\boldsymbol{\Delta}\mathbf{c}^{(i)} = 2\ {\mathbf{W}^{(i)}}^{-1}.\mathbf{R}^{(i-1)}, \tag{4.25}$$

where we define the $(2i-1)$ vectors $\boldsymbol{\Delta}\mathbf{c}^{(i)}$ and $\mathbf{R}^{(i-1)}$, and the matrix $\mathbf{W}^{(i)}$ of size $(2i-1)\times(2i-1)$ as follows:

$$\boldsymbol{\Delta}\mathbf{c}^{(i)} = \begin{pmatrix} \Re(\Delta c^{(i)}_{\boldsymbol{\vartheta}_1}) \\ \vdots \\ \Re(\Delta c^{(i)}_{\boldsymbol{\vartheta}_i}) \\ \Im(\Delta c^{(i)}_{\boldsymbol{\vartheta}_2}) \\ \vdots \\ \Im(\Delta c^{(i)}_{\boldsymbol{\vartheta}_i}) \end{pmatrix} \qquad \mathbf{R}^{(i-1)} = \begin{pmatrix} \Re(R^{(i-1)}_{\boldsymbol{\vartheta}_1}) \\ \vdots \\ \Re(R^{(i-1)}_{\boldsymbol{\vartheta}_i}) \\ \Im(R^{(i-1)}_{\boldsymbol{\vartheta}_2}) \\ \vdots \\ \Im(R^{(i-1)}_{\boldsymbol{\vartheta}_i}) \end{pmatrix} \tag{4.26}$$

$$\mathbf{W}^{(i)} = \begin{pmatrix} \mathbf{W}^{(i)}_{11} & \mathbf{W}^{(i)}_{12} \\ \mathbf{W}^{(i)}_{21} & \mathbf{W}^{(i)}_{22} \end{pmatrix}$$

with:

$$\begin{aligned}
\mathbf{W}^{(i)}_{11} &= \left[\Re(W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y} + W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y})\right]_{(x,y)\in[\![1;i]\!]\times[\![1;i]\!]} \\
\mathbf{W}^{(i)}_{12} &= \left[\Im(W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y} - W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y})\right]_{(x,y)\in[\![1;i]\!]\times[\![2;i]\!]} \\
\mathbf{W}^{(i)}_{21} &= \left[\Im(W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y} + W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y})\right]_{(x,y)\in[\![2;i]\!]\times[\![1;i]\!]} \\
\mathbf{W}^{(i)}_{22} &= \left[\Re(W_{\boldsymbol{\vartheta}_x-\boldsymbol{\vartheta}_y} - W_{\boldsymbol{\vartheta}_x+\boldsymbol{\vartheta}_y})\right]_{(x,y)\in[\![2;i]\!]\times[\![2;i]\!]}
\end{aligned}$$

Consistently with these definitions, we conventionally force the first selected frequency $\boldsymbol{\vartheta}_1$ to be the zero frequency. Since the signals considered in our work are real-valued, so are their average value on the hyper-block. The corresponding expansion coefficients have therefore no imaginary part, and the corresponding null rows/columns are removed from the matrix definitions. Hence the $(2i-1)$ size instead of the expected $2i$.

Eventually, the parametric model and the residue are updated in the Fourier domain according to all the i selected basis functions. Thus, for every $\boldsymbol{\vartheta} \in \Theta^{(i)}$, we have:

$$\begin{cases} G^{(i)}_{\boldsymbol{\vartheta}} = G^{(i-1)}_{\boldsymbol{\vartheta}} + \frac{1}{2}\ P.\Delta c^{(i)}_{\boldsymbol{\vartheta}} \\ G^{(i)}_{-\boldsymbol{\vartheta}} = G^{(i-1)}_{-\boldsymbol{\vartheta}} + \frac{1}{2}\ P.\Delta c^{*(i)}_{\boldsymbol{\vartheta}} \end{cases} \tag{4.27}$$

$$R^{(i)}_{\boldsymbol{\vartheta}} = R^{(i-1)}_{\boldsymbol{\vartheta}} - \sum_{\boldsymbol{\vartheta'}\in\Theta^{(i)}} \frac{1}{2}\left(\Delta c^{(i)}_{\boldsymbol{\vartheta'}}.W^{*}_{\boldsymbol{\vartheta'}-\boldsymbol{\vartheta}} + \Delta c^{(i)*}_{\boldsymbol{\vartheta'}}.W_{\boldsymbol{\vartheta'}+\boldsymbol{\vartheta}}\right) \tag{4.28}$$

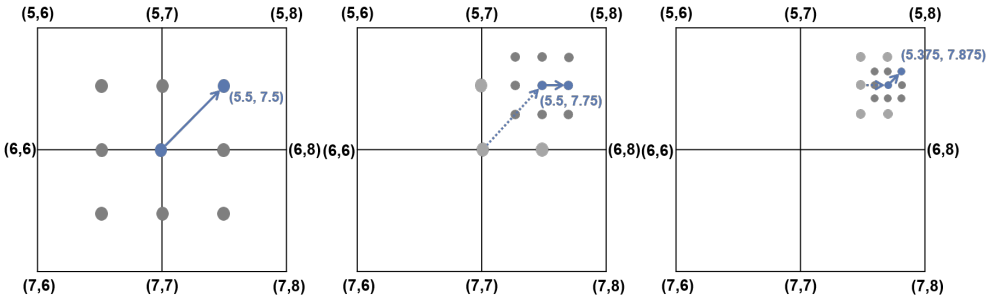


Figure 4.4: Refinement example. Shifting the integer frequency by a small step to one of the eight directions at each iteration/refinement level.

4.2.5 Frequency Refinement to Non-integer Values

So far, the approximation model is generated by including the discrete Fourier functions that best represent the available data samples. However, the actual spectrum of the light field signal is not necessarily aligned with the discrete sampling grid. This is why windowing effects may be observed in light fields DFT 4D-spectra.

Besides, since the angular resolution of a light field (its number of views, or of sub-aperture images) is usually lower than its spatial resolution (the number of pixel per view or per sub-aperture image), the windowing effect is stronger in the angular dimensions, as observed in [22]. Indeed, Shi *et al.* [22] proved that the sparsity of the light field is better preserved in the continuous Fourier domain, thus the need to optimize the estimated frequency positions into non-discrete values, in order to enhance the approximation of the light field Fourier spectrum.

We propose to refine the discrete angular frequencies to non-integer values to approach the continuous Fourier spectrum of the light field, and thus recover its sparsity level. The refinement can also be performed in the spatial dimensions, but with a lesser impact since light fields usually present much higher spatial than angular sampling.

Unlike the approach of Shi *et al.* [22], where the recovery of continuous frequencies is conducted in the full light-field spectrum, as a global post-processing, the Orthogonal Frequency Selection-related refinement is performed locally, at the hyper-block scale, directly within the loop of frequency selection.

Each time a new frequency is selected via residual energy minimization, we shift its position by a small fractional step δ to all the eight angular directions shown in Figure 4.4. The residual energy is calculated in all the eight corresponding frequency positions. The position that maximizes the residue decrease is selected as the final frequency to include to the model. An overview of the refinement is detailed in Algorithm. 1.

The reconstruction steps that follow the selection are the same as in the regular (discrete) case, except that computations can no longer be performed in the DFT domain, but in the pixel domain.

Algorithm 1 Angular Frequency refinement

$\Delta = \{(-1,-1),(-1,0),(-1,1),(0,-1),(0,1),(1,-1),(1,0),(1,1)\}$
$\delta = 1/2$
define max_ref
Select a frequency position $\boldsymbol{\vartheta} = (\mu, \nu, \zeta, \xi)$
$largest\_R = |R(\boldsymbol{\vartheta})|$
$i_{ref} = 1$
while $i_{ref} \leq max\_ref$ **do**
    for $(d\zeta, d\xi) \in \Delta$ **do**
        $\boldsymbol{\vartheta'} = \boldsymbol{\vartheta} + (0, 0, \delta.d\zeta, \delta.d\xi)$
        Calculate $R(\boldsymbol{\vartheta'})$
    $\boldsymbol{\vartheta^*} = \text{argmax}_{\boldsymbol{\vartheta'}} |R(\boldsymbol{\vartheta'})|$
    if $|R(\boldsymbol{\vartheta^*})| > largest\_R$ **then**
        $\boldsymbol{\vartheta} = \boldsymbol{\vartheta^*}$
        $largest\_R = |R(\boldsymbol{\vartheta^*})|$
    $\delta = \delta/2$
    $i_{ref} = i_{ref} + 1$

There are several advantages in refining frequencies as soon as they are selected, rather than as a hyper-block post-processing. On-the-fly refinement helps converging swiftly to the actual light field spectrum, instead of dealing with several discrete frequencies that correspond to the same spectrum peak. Thus, the number of iterations is reduced, and the overall reconstruction quality is improved.

4.3 Experimental Setup and Results

In order to evaluate the proposed approach, we apply it to various synthetic and real light field datasets: plenoptic images [48], multi-view contents[1]. We evaluate the reconstruction quality by PSNR and SSIM measurements for different sampling rates.

4.3.1 Parameter Settings

In order to keep the computational load manageable, the transform model size is set to 32x32x5x5, which means a block size of 4x4x3x3, and spatial and angular border widths of 14 and 1 respectively.
As for the weighting function parameters, we set the decay factor ρ_s in spatial direction to 0.7 and ρ_a in angular direction to 0.5. This ensures a higher contribution of spatial neighbors compared to angular neighbors, following the sparsity structure of the light field. Reconstructed areas will have a weight of $\sigma = 0.5$ compared to originally available samples. Since the maximum number of frequencies a hyper-block can contain is equal to its size (here 144), and that at each iteration the OFS method permits to add 2 frequencies (the selected one and its conjugate), we set the number of iterations at 72. To make a fair comparison, we set this parameter to 200 for the non orthogonal FSR method, where the algorithm could select more than 144 frequencies, with some duplication. Finally, for the refinement step, we decided to limit the maximum level of refinement to $max\_ref = 3$.

4.3.2 Results

We show in this section experimental results on various light fields of the Frequency Selection-based methods FSR and OFS. The proposed reconstruction algorithm is compared against several state-of-the-art methods: Miandji *et al.* [125] using trained dictionaries to reconstruct light fields in a compressive sensing framework, Shi *et al.* [22] that use the same assumption that light fields are sparse in the Fourier domain, and Kalantari *et al.* [48] which synthesize all light field views from a subset of views using a deep learning-based framework.

Synthetic dataset

We first present a comparison with a compressive sensing-based method [125] that uses learned sparse dictionaries to reconstruct the light field via overlapping or non-overlapping 2D patches. The number of views of the tested light field *Dragon*[2] is 5x5. The input data for both methods is the same randomly sampled light field with various predefined sampling rates. The PSNR results are summarized in Figure 4.5. As shown in the graph, our OFS method achieved a much higher quality than in [125], even for a very low sampling rate, with an average PSNR gain of 4.43dB. The non-integer refinement also adds an average PSNR gain of 1.2dB to the OFS method.

In Figure 4.6, we show the reconstructed central views and the differences from the ground truth of *Dragon* at a sampling rate of 4%. One can see that a better visual quality is achieved by the OFS method over the dictionary-learning method.

[1] http://lightfield.stanford.edu/lfs.html
[2] http://web.media.mit.edu/ gordonw/SyntheticLightFields

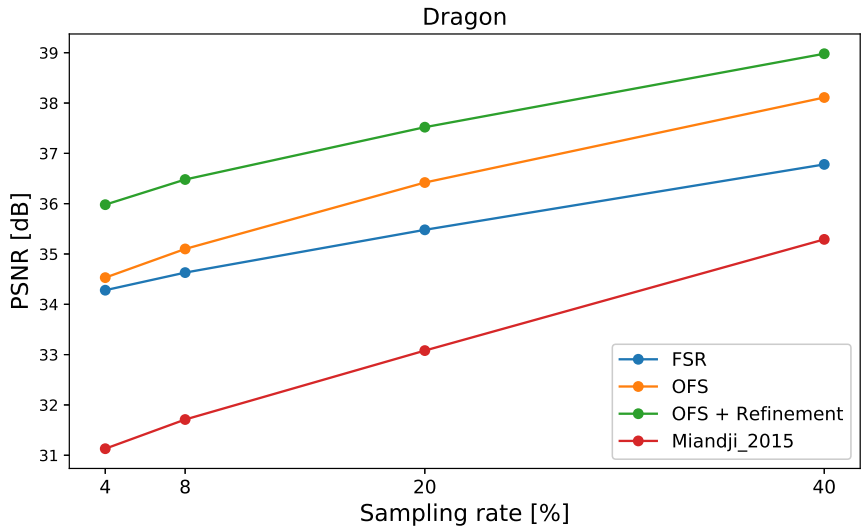


Figure 4.5: PSNR comparison of 4D-FSR, 4D-OFS and 4D-OFS with refinement, with a state-of-the-art method: Miandji *et al.* [125] for different sampling rates.

Real Light Field dataset

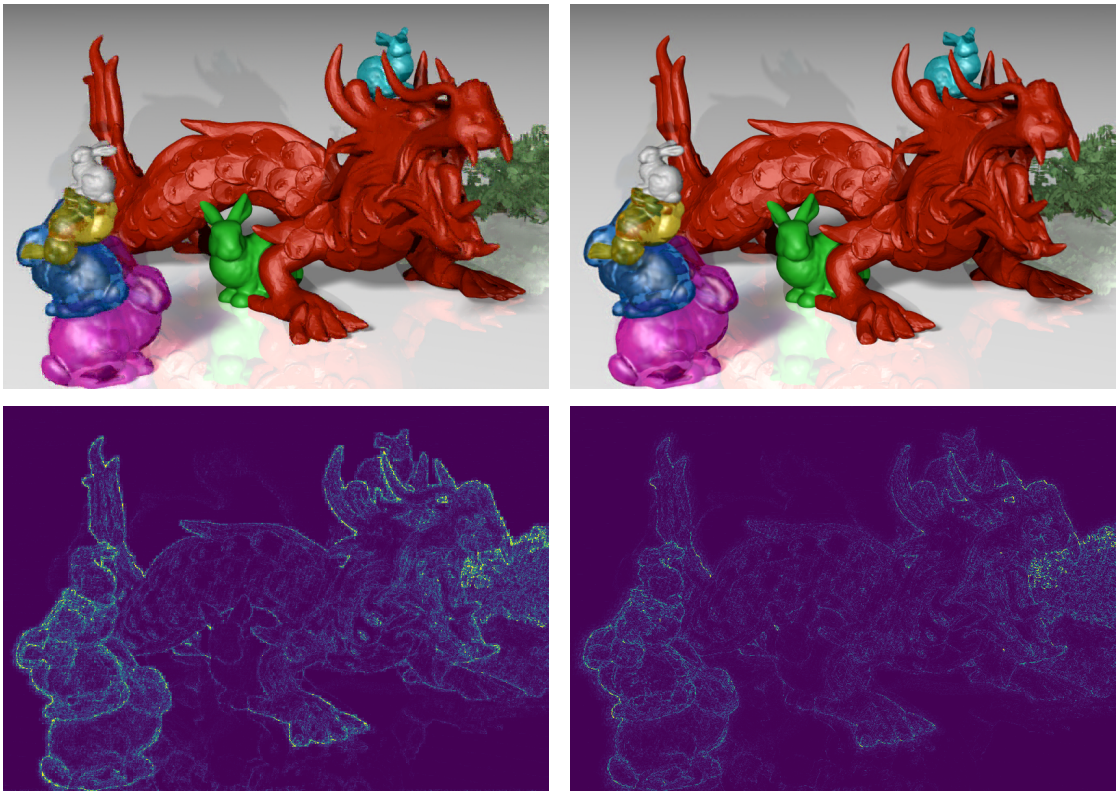
We next show another comparison with a state-of-the-art method [22] that uses the Sparse Fourier Transform (SFFT) to reconstruct a light field from a subset of views, spanning over 1D viewpoint trajectories.
The tested datasets are taken from the Stanford light field Archive[3]. We reconstruct each light field from a sampled set of 81 views. The method [22] has been tested with a number of 45 input views chosen using their box-and-X pattern, which corresponds to $45/81 = 0.55$ of the full light field. We use this same ratio for our input data sampling.

Table 4.2 presents the reconstruction quality of both methods on various light fields: *Bunny*, *Crystal*, *Amethyst* and *Lego Knights*. The PSNR and SSIM values are averaged over the 9x9 reconstructed views. One can see that our method achieves a much better reconstruction quality with a PSNR average gain of 7.37dB.

Besides, Figure 4.7 presents reconstruction examples on *Crystal* and *Amethyst* corresponding to the viewpoint $(5, 2)$. The difference images show a very strong unstructured noise in the SFFT method result. This noise cannot be related to the original data since we do not find it in our reconstruction result, but is more likely to be inherent to the initialization step in [22] where a rough estimation of the frequency positions is made using a voting strategy from the available views spectra.

In the highly non-Lambertian scene *Amethyst*, our method succeeds in reconstructing the specular and reflective features, with very small difference to the original data, while the SFFT-based method [22] fails in reconstructing some zones of reflections/refraction.

[3] http://lightfield.stanford.edu/lfs.html



Miandji *et al.* [125] Ours

Figure 4.6: Reconstruction quality comparison on the light field *Dragon* at 4% sampling rate. Top: reconstructed images. Bottom: difference from the ground truth (magnified by 5).

Finally, we conducted experiments on plenoptic light field data sets from [48]: it consists of 8x8-view light fields. We compare our method to three deep learning-based reconstruction methods: Kalantari *et al.* [48] use 4 corner viewpoints to reconstruct the full 8x8 views, Vadathya *et al.* [128] reconstruct the light field at 7x7 angular resolution from a coded image, and finally Nabati *et al.* [129] use a coded color mask to even lower the sampling rate to $1/(25\text{x}3) = 1.3\%$ to reconstruct 5x5 viewpoints.

We apply the same compression ratios to our method for comparison on different light fields: *Flower 1*, *Rock*, *Flower 2*, *Seahorse* and *Cars*. The results are summarized in Figure 4.8. The PSNR results for [128] and [129] were extracted from the corresponding articles respectively, more results could be presented for the method in [48] for which the software was available.

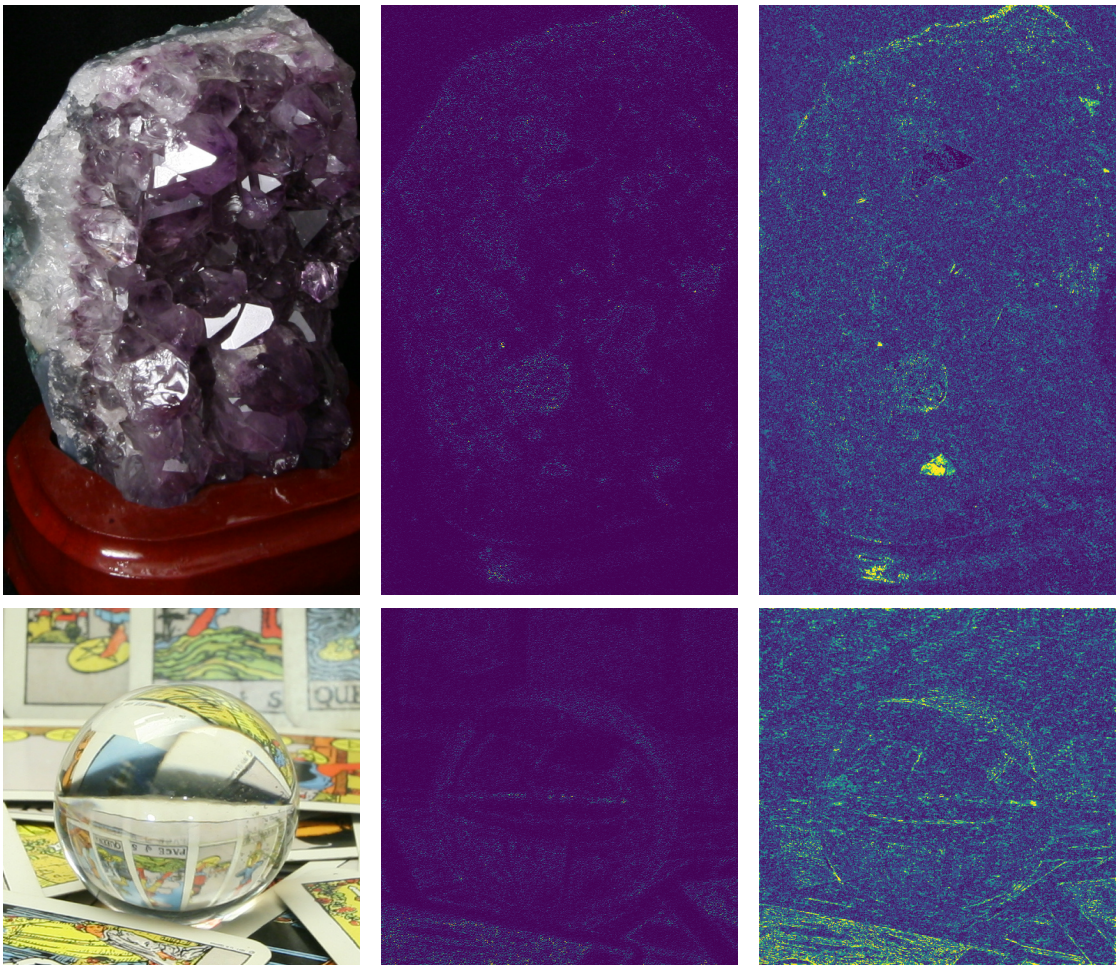
As shown in the graphs of the figure, our method outperforms the approach in [48] of at least 1.22 dB, and achieves high PSNR values of more than 32dB at only 5% of input data. The reconstruction quality is lower nevertheless if we take a very low number of samples, of 2% or lower, where the compressive sensing theory is limited to achieve a correct restoration of the

Table 4.2: Reconstruction quality comparison of real light fields from the Stanford Gantry dataset.

	Shi *et al.*		Ours	
	PSNR(dB)	SSIM	PSNR(dB)	SSIM
Bunny	40.44	0.9797	**47.49**	**0.9969**
Crystal	32.96	0.9633	**41.04**	**0.9956**
Amethyst	35.33	0.9379	**41.91**	**0.9906**
Lego Knights	31.85	0.8277	**39.65**	**0.9763**
Lego Truck	39.18	0.9410	**43.75**	**0.9860**

missing data.

Table 4.3 shows a visual comparison of our reconstruction results to the ones from [48] on the view $(4, 4)$ of each 8x8 tested light field. The difference images demonstrate some diffuse small errors in our results with a minimum value of SSIM equal to 0.969 while the results in [48] exhibit important errors on edges, probably related to the disparity estimation limitations.



Ground truth Ours Shi *et al.* [22]

Figure 4.7: Reconstruction quality comparison with Shi *et al.* [22]. Top: Amethyst. Bottom: Crystal (difference images are magnified by 10).

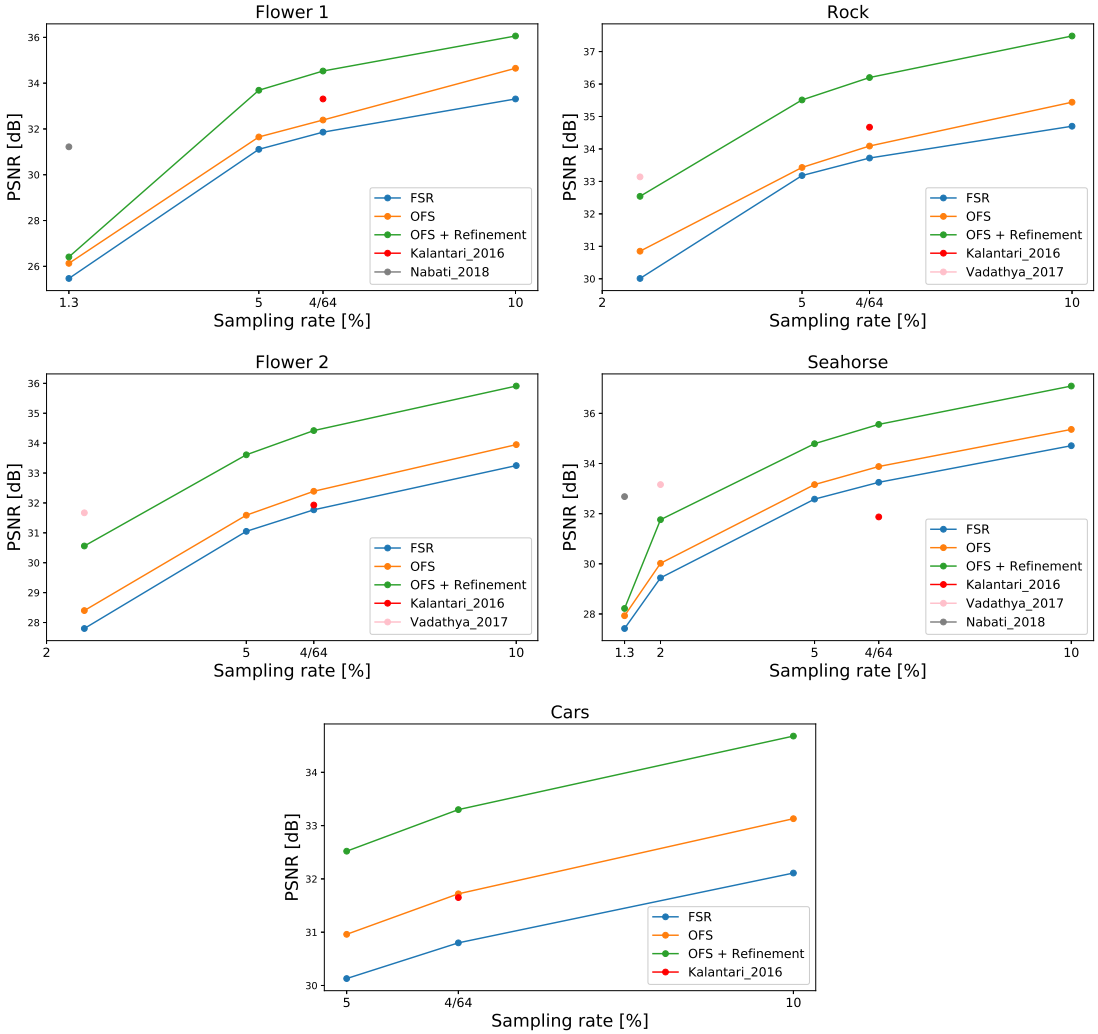


Figure 4.8: PSNR of light fields reconstructed with different methods: Kalantari *et al.* [48], Vadathya *et al.* [128], Nabati *et al.* [129], FSR [1] in 4D, and ours (OFS and OFS+refinement).

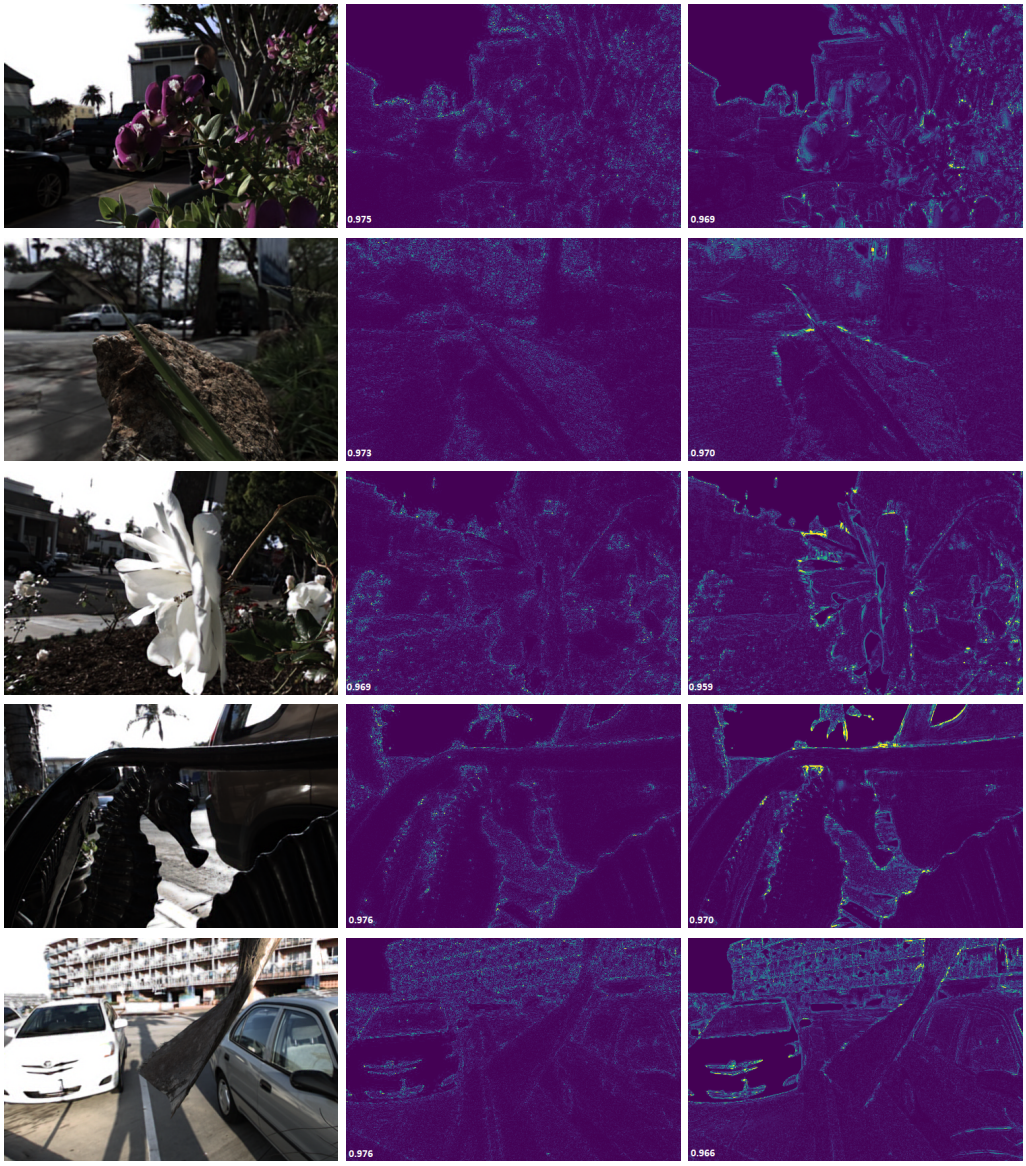


Table 4.3: Evaluation of the reconstruction quality of different methods. Light fields (Top to bottom) *Flower 1*, *Rock*, *Flower 2*, *Seahorse* and *Cars*. (Left to right) our reconstruction image, difference of our result to ground truth, difference of result from [48] to ground truth (difference images are magnified by 5).

4.4 Conclusion

In this chapter, we introduced a new iterative block-wise algorithm to compressively reconstruct light field images. We tackled the challenge of capturing high-resolution light field images of the scene, by storing compressive data and reconstructing the full resolution images using an approximation of the available samples in the Fourier domain. The approximation model is generated by sparsely selecting the Fourier basis functions that best fit the sampled data, while ensuring the orthogonality of the residue to the subspace spanned by the already selected basis functions. The angular frequency positions are furthermore refined to non-integer values in order to preserve the sparsity that may be limited by the small angular sampling.

Experimental results show that high-quality reconstruction is achieved by our approximation method, and demonstrate the advantage of the improved version, which increases the quality of the reconstruction while using a reduced number of iterations. Moreover, refining the model to non-integer frequency positions permits a better approximation. Comparisons with state-of-the-art methods demonstrate that our approach is competitive both in terms of PSNR and SSIM, even for low sampling rates.

More importantly, our solution does not require multiple shots, or any prior knowledge of the captured scene geometry. Therefore, the proposed method can be extended to applications that require real-time shooting, such as very high-resolution light field videos.

Chapter 5

Towards an End-to-end Light Field Image System

5.1 Introduction

One of the major research concerns in light field imaging is the data quality conservation through the various applications. Indeed, many attempts to adapt several image processing to the inherent information that light fields present have demonstrated certain limitations in the visual quality of the resulting data, either related to some detail information loss (*e.g.* due to occlusions, reflectance...) or ineffective methods of depth estimation, or inadaptability of coding schemes to the light field data structure. Thus, these quality issues make it difficult to use the resulting light field data for high-end image and video products and applications.

With the ever-growing need to high quality image and video content, such as for virtual reality applications, it is of high interest to explore novel possible processing approaches and especially evaluate their global impact on the final produced light field.

The previous chapters are related to two main aspects of light field processing, and proposed solutions for efficient compression and reconstruction of light fields on both the acquisition and transmission sides.
In order to evaluate the impact of the several processing steps on the quality of the final light field, we propose to study the full end-to-end scheme from the capture to the broadcast of light field data. We measure the quality variation through tests on several light fields, and conclude on the coding performance that our scheme could provide under low compression ratios.
To our knowledge, there is only one work that proposes a full system of acquiring, processing and compressing light fields [153]: the system is composed first of a rig of GoPro cameras for acquisition of panoramic light fields. A compression approach is introduced by modifying the VP9 video codec for an efficient compression performance with real-time, random access and decompression.

Most light field HEVC-based compression schemes either encode the full light field grid of views in a pseudo-sequence arrangement, or directly encode the lenslet image. Many compressive capture methods have proven to be efficient to recover the full-resolution light field from a set of samples (masked sub-aperture images, a subset of views...), but no further study on their impact for light field transmission has been done, especially for low sampling rates. Indeed, although the quality of the reconstructed light field is good, there is no guarantee on how the transmission encoder could maintain this quality.
Through the study below, we propose to assess the impact of our compressive reconstruction method on the quality of the output light field of our scalable coding scheme introduced in Chapter 3.

5.2 Overview of the Experimental Study

We propose to run experimental tests on the full light field processing scheme, composed first of a compressive sensor that captures randomly sampled light fields using the scheme explained in Chapter 4, followed by a reconstruction step based on the Orthogonal Frequency Selection (OFS) where the 4D spectrum of the captured light field is recovered and a full-resolution light field signal is generated. The output light field will be compressed using our scalable coding scheme introduced in Chapter 3 for the transmission step. Figure 5.1 shows an overview of the full acquisition and compression scheme.

The input of the full scheme is a 4D light field sampled at a pre-defined sampling rate, and the output is a 4D light field resulting from the reconstruction of the missing data samples and their encoding-decoding. The objective here is to evaluate the distortion that our scheme introduces to the original light field data.

Several experiments are conducted on different light field datasets to study the impact of the compressive reconstruction and compression chain on the quality of the output data. Optimal compression parameters could be then deduced from the quality evolution in the rate-distortion results.
Real plenoptic images from the plenoptic JPEG Pleno Dataset [149], as well as multi-view light fields from the Stanford Gantry Dataset [1] have been used for the tests.

Different compression ratios have been tested: first, for the compressive acquisition, sampling rates ranging from 5% to 40% have been used to assess the reconstruction performance and its impact on further processing. Second, the coding was performed using the following settings for the Quantization Parameter (QP): 22, 27, 32 and 37, to illustrate the quality variation with the global compression rates. The objective quality is evaluated on the YUV components with PSNR averaged on all the views of the light field. The bit-rate is calculated from the coded bitstream for all YUV components. The light field images are encoded into YUV sequences with a frame rate fixed at 50 fps.

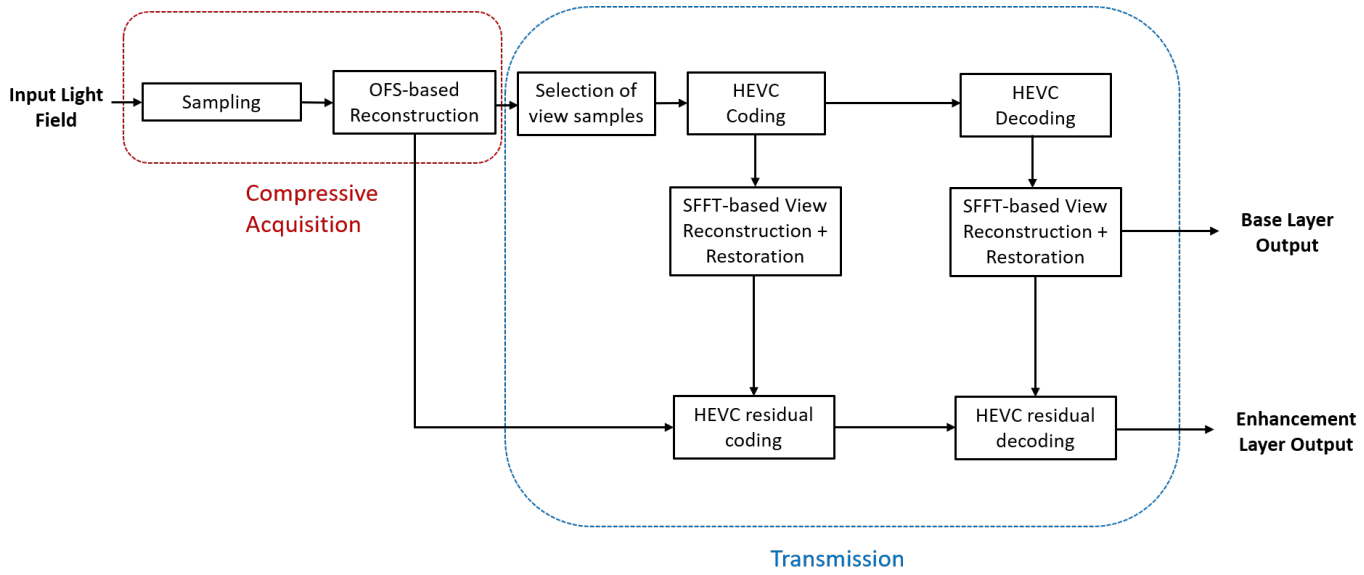


Figure 5.1: Overview of the full scheme of compressive acquisition and transmission.

[1] http://lightfield.stanford.edu/lfs.html

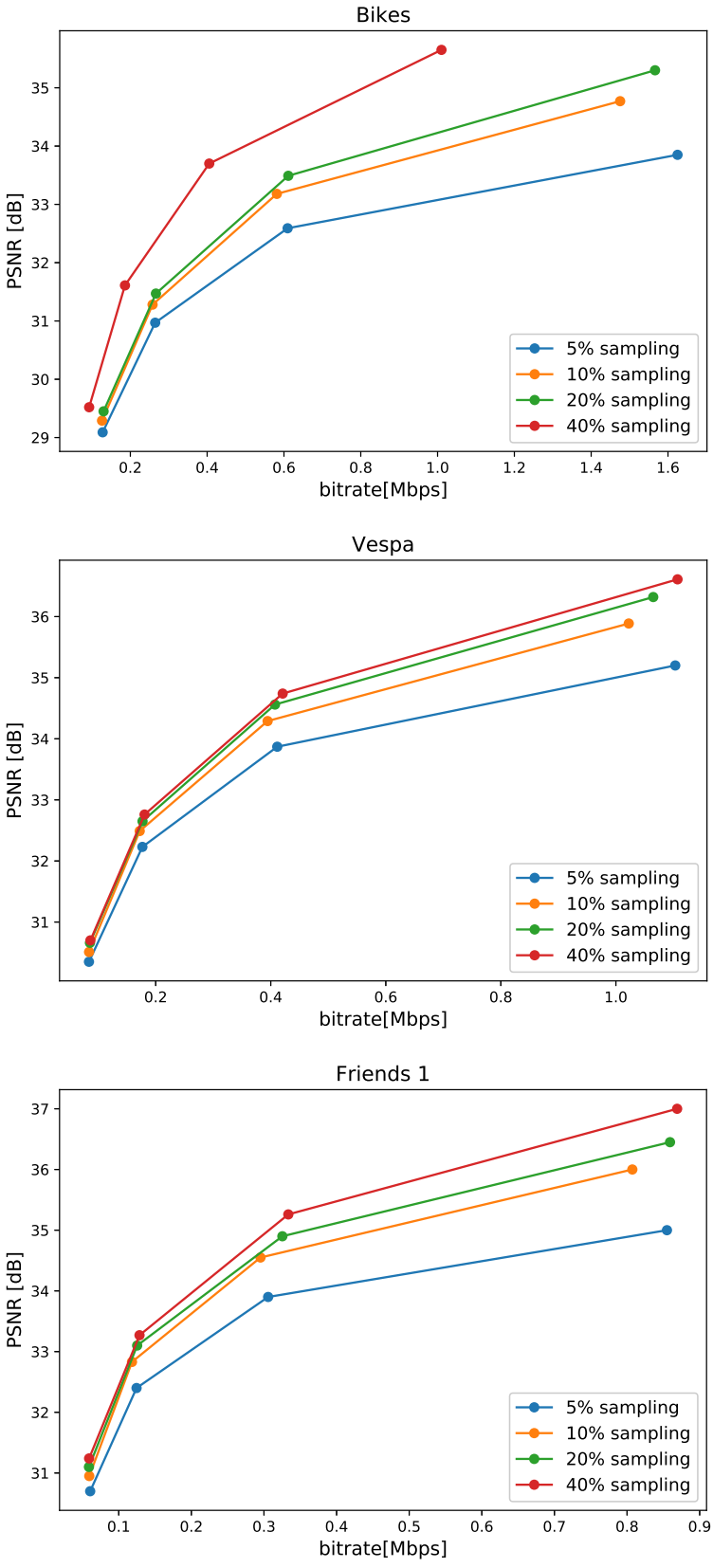


Figure 5.2: Rate-distortion results for plenoptic light fields from the JPEG Pleno Dataset [149].

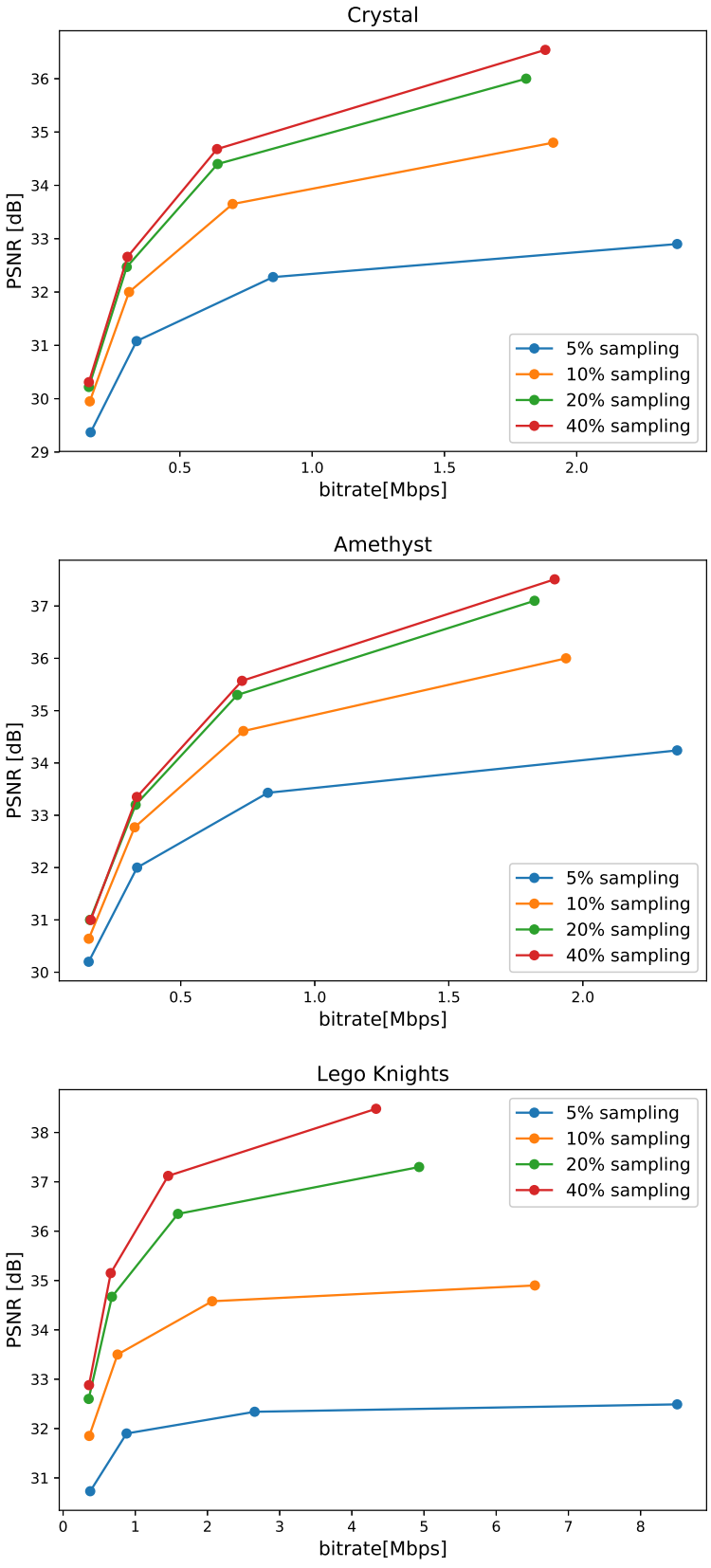
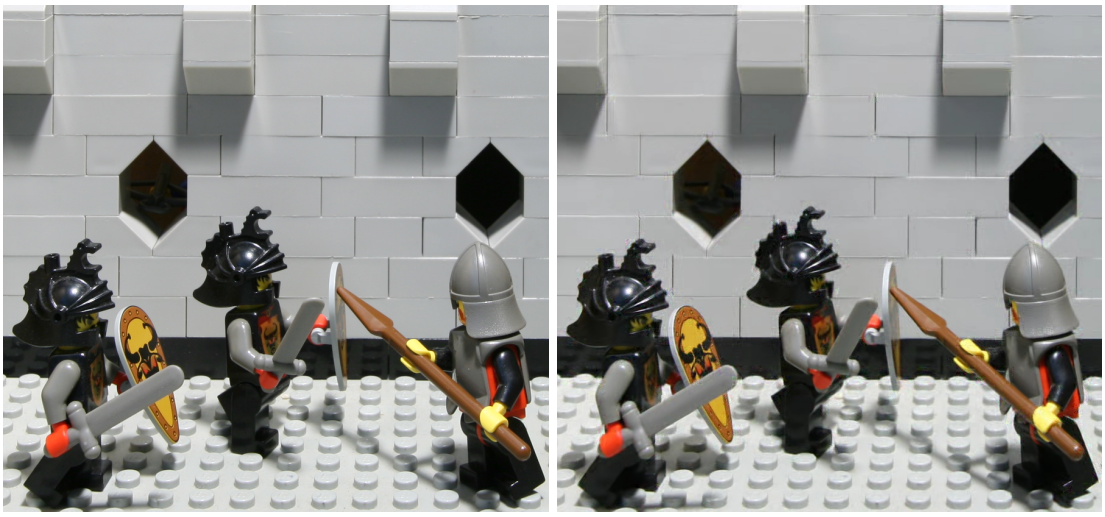


Figure 5.3: Rate-distortion results for Gantry-based light fields from the Stanford Dataset.

5.3 Results

From the tests described above, and in order to represent the evolution of the distortion caused by our acquisition to transmission scheme, rate-distortion curves are calculated and presented in Figure 5.2 and Figure 5.3. Each curve represents the PSNR variation with the coding bitrate (Mbps) and corresponds to one sampling rate in the compressive sensing step.
The PSNR results show that using the full light field compressive scheme globally provides a good quality level of the decoded light field. The quality coherently varies with both the quantization and sampling levels: indeed, it increases with the number of available input samples, and is inversely proportional to the Quantization Parameter.

Besides, even for very low sampling rates on the acquisition side, light fields are restored with a fair quality reconstruction. However, we observe that the PSNR-rate performance saturates for small sampling rate (5%). This phenomenon is especially apparent in Rate distortion curves of the multi-view light field datasets (Crystal, Amethyst and Lego Knights) in Figure 5.3. The RD-performance for the light field *Lego Knights* is more specifically limited in high bit-rates. This can be related to the structures that images of this light field contain, and that are source of prediction errors near edges (see Figure 5.4).



Original central view Decoded central view

Figure 5.4: Result example of our full scheme. The decoded image corresponds to QP=22 and a sampling rate at 5%. One can observe some artefacts on edges of the structures in the background.

Comparison to a DIBR-based coding Approach

We propose in this section to compare the coding results of our scheme to a depth-based approach [154]. This method encodes 4 corner views of the light field in a pseudo-sequence using

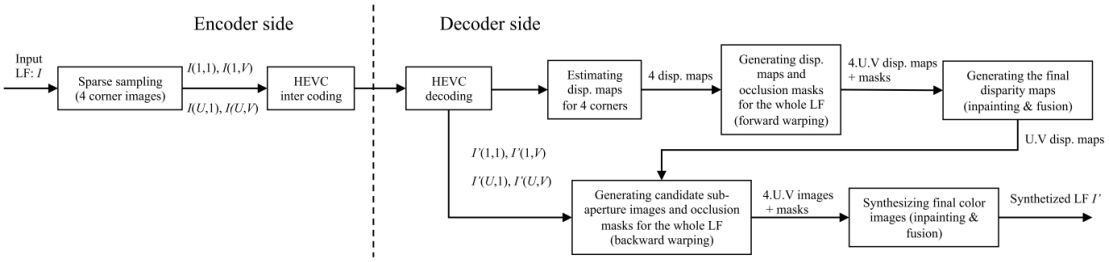


Figure 5.5: Overview of the coding and decoding scheme from [154].

HEVC inter coding, and synthesizes the remaining views on the decoder side by estimating the disparity maps (see Figure 5.5). The use of only 4 input views reminds of a compressive acquisition of a higher-resolution of a light field.

The disparity map of each corner view is calculated using DeepFlow [155], and the disparity maps of the synthesized views are projected from these reference maps. The missing information is recovered by solving a low rank approximation problem, taking advantage of the high correlation between the views. The color views are then synthesized by warping the available views and the same information recovery approach, as for the disparity maps, is used to fill the holes in the warped images.

The work in [154] proposed to compare its disparity map-based synthesis to another CNN-based view synthesis approach [48], introduced in the decoder side to generate all the light field views. Furthermore, to improve the quality of the final light field, the residual of the whole light field is encoded as a pseudo-video sequence by HEVC inter coding. This is similar to our proposed enhancement layer coding, in which the residue of reconstructed light field in the base layer is used.

The three tested light fields in the work [154] are from the Lytro Illum light field dataset of Irisa [2]. The images *Building*, *Fruits* and *Rose* contain 8x8 views of 541x376 pixels each. The central views are shown in Figure 5.6. The corresponding sub-aperture images are extracted (*i.e.* demultiplexed) from RAW capture using the *Lytro Power Tools Beta* software [3].

To make a fair comparison to this method, we first sample the light field by $4/64 = 6.25\%$ and recover it using the OFS-based reconstruction method. The QP values are chosen similarly to those used in the DIBR-based approach [154]: 14, 20, 26 and 32. PSNR values are computed on the luminance component Y.

We first compare the single layer coding approaches using view synthesis (using DIBR and CNNs) with our scheme considering only the base-layer output data (Compressive acquisition + single layer coding). We then compare the extended method that uses a residual coding from [154] to our full scheme containing the scalable coding for transmission (Compressive acquisition + scalable coding).

[2] https://www.irisa.fr/temics/demos/IllumDatasetLF/index.html

[3] http://lightfield-forum.com/lytro/lytro-archive/lytro-platform-lytro-power-tools-lytro-development-kit-official-product-information

Figure 5.6: Light Fields used in the tests: *Building*, *Fruits* and *Rose*.

The rate-distortion curves in Figure 5.7 present the light field quality results from the different approaches: one can observe that our scheme using one single coding layer outperforms the view synthesis-based coding approaches for the light fields *Building* and *Fruits*, but not for *Rose* which exhibits a uniform texture and a small disparity variation, which could explain the advantage of the depth map-based view synthesis [154].

However, both the PSNR values resulting from our approach and the disparity-based methods reach a plateau starting from a certain bit-rate level. Indeed, due to the small sampling rate (6.25%), the quality of the light field at the end of the compression chain cannot go beyond the maximum quality that the sparse OFS-based reconstruction provides after the compressed acquisition.

In the case of the disparity-based scheme for light field compression presented in [154], the quality of the estimated disparity maps, which depends both on the intrinsic performances of

the DeepFlow estimator and the quality level of the input images, presents a limiting factor for the accuracy of the synthesized final images. The small number of input views limits also the performance of the depth estimation of all the views, since a large parallax needs to be recovered. The higher the inter-view parallax is, the less accurate the disparity estimation is. Since the plenoptic images used here for the tests exhibit small disparities, one can imagine more difficulties for these disparity-based methods to accurately recover light fields with wide baselines.

Besides, the added residual coding (in green line) of the DIBR-based method in [154] provides an enhancement to the compression performance of the single layer coding version (in blue line). The residue coding is generated using a fixed QP for the base layer (of the 4 corner views), and varying QP values for encoding the residue of all the views. Our full scheme using scalable coding outperforms the DIBR scalable method for low bit-rates for the datasets *Building* and *Fruits* but presents lower performance in higher bit-rates. Indeed, the method [154] requires lower bit-rates to encode the 4 corner views, but adds complexity in the decoded side with the view synthesis of the remaining views. Moreover, the QP value for their base layer [154] is fixed to 20 and only the QPs residual coding varies, while in our scheme, both base-layer and enhancement-layer QPs vary.
The global performance results validate the interest of using our full compressive acquisition and transmission scheme, and imply promising improvement possibilities for further exploitation of light field sparsity in the design of compression solutions.

5.4 Conclusion

We explored in this chapter the possibility to include the two previously introduced approaches to compose a complete light field processing chain in which we study the compressibility of light field contents. The out-coming image quality demonstrates the global efficiency of the scheme in terms of compression and recovery of light field images. The results demonstrate the capability of our compressive scheme to recover several light fields with a fair quality level, and show the advantage of our approach to depth-based approaches [154] in low bit-rates.
Indeed, our scheme does not rely on depth information but extends signal processing techniques to reconstruct the light field by taking advantage of its sparsity, it still offers good RD-performance comparable to the depth-based reconstruction methods.

Further research could be conducted in order to optimize the full processing scheme and also to adapt it to light field videos. First, the sample selection for the input of the base-layer codec can be better adapted to the acquisition sampling rate in order to ensure higher quality while decreasing the coding cost. A trade-off should be found between the parameters of the compressive capture and transmission. Moreover, a learning-based study may help calculating the optimal compression parameters for the full scheme. Finally, an improved method for recovering the missing views in the base layer could also be proposed to overcome the existing limitations of the SFFT-based method.

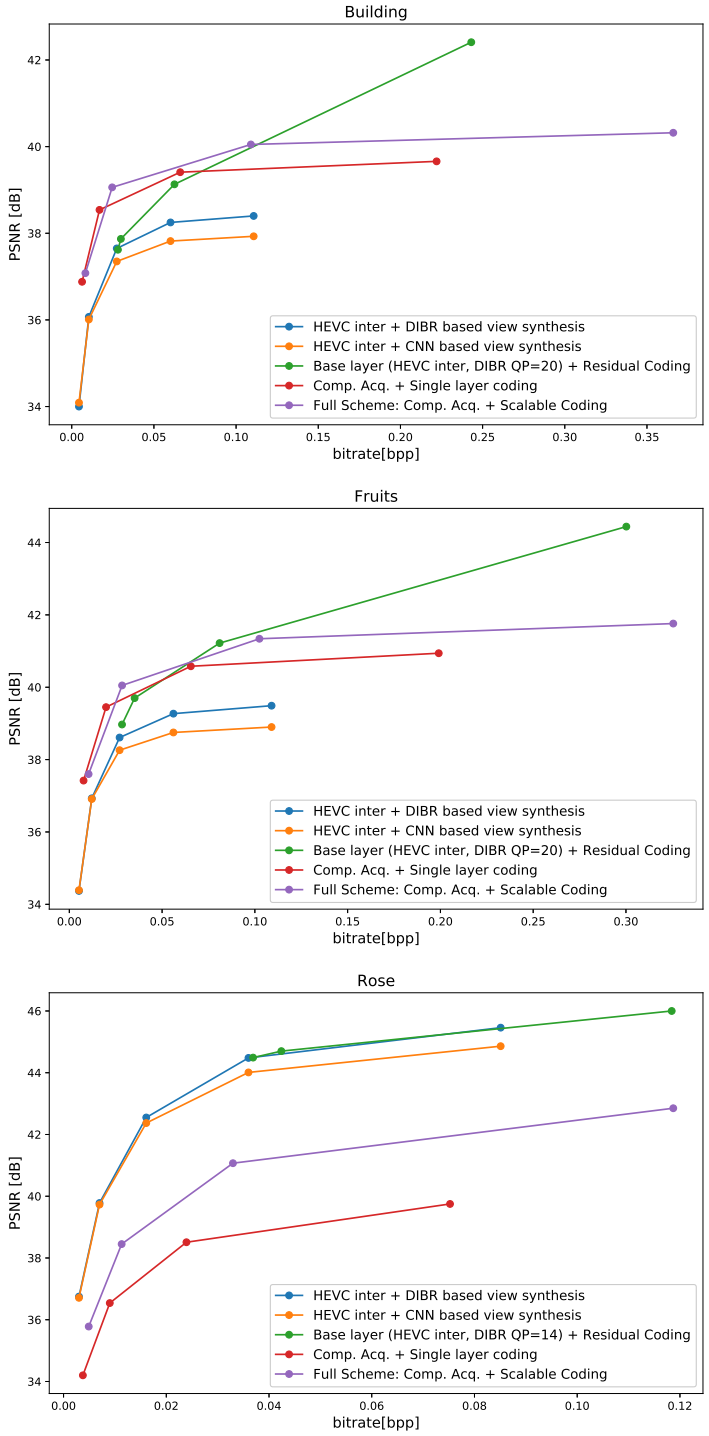


Figure 5.7: Comparison of rate-distortion results for plenoptic images with the depth-based method [154].

Conclusion

The constant progress of new image technologies and devices that provide more depth information of the captured scene has produced huge volume of data that require more efficient compression solutions than the classical coding schemes. A fundamental step in the conception of new compression algorithms is to find the optimal representation, and more specifically, the sparse representation that yields the use (or the capture) of the least number of images, while producing a high-resolution light field content. In this thesis, we were interested in studying the compressibility of light field data, and more particularly the use of light field sparsity to provide a more efficient compression performance than standard coding schemes, but also to make possible the compressive acquisition of such content in order to capture higher sampling of the light field.

We first targeted the problem of light field transmission. We proposed a scalable compression scheme that reduces the number of input images of the encoder and enables the reconstruction of the whole light field by a sparse Fourier-based reconstruction algorithm. The sparsity is described in the angular dimension to recover the missing angular samples (or sub-aperture images) of the light field. The proposed method has proven to be more efficient than the standard HEVC inter coding of all the light field. It also achieves high compression performance for contents with a large baseline such as moving gantry camera-based light fields, providing an average bit-rate reduction of 11.2%. The study of the impact of compression from our scheme on the generation of all-in-focus images have confirmed the advantage of our method compared to the baseline method.

Another important aspect of the light field imaging is the capturing system. In order to provide high-resolution light fields, the existing acquisition devices capture multiple images from different viewpoints of the scene. The number of images is inherent to the number of cameras (in the case of camera-array acquisition) or the micro-lenses (in the case of plenoptic cameras), and their resolution being limited as well, the output light field still suffers from a spatio-angular trade-off due to these technical limitations. To overcome these constraints, the light field can be compressively acquired, by only capturing and storing a small number of spatio-angular samples, and a resampling can be performed to generate a high resolution version. The sparsely selected samples can accurately describe the scene within the light field and allow a complete reconstruction of the full-resolution light field.

In Chapter 4, we proposed a compressive reconstruction method of randomly sampled light field data using sparse representation in the Fourier domain. The method iteratively selects the frequencies that best approximate the Fourier transform of the original light field from the available input samples. The frequency selection is made optimal by ensuring an orthogonality

with the already selected frequencies, so that the residual error decreases faster and less alteration of the frequency weights is caused. From the observation that sparsity is better conserved in the continuous spectrum domain, another improvement was introduced that refines the initially discrete frequency positions to non-integer values, in order to further approximate the continuous Fourier spectrum of the original light field.

In comparison with several state-of-the-art approaches based on sparse models or deep learning reconstruction, our method achieves high-quality reconstruction of various light field contents. Even for very low sampling rates of around 5%, the reconstructed light fields have a quality level in PSNR over 34dB. The proposed approach is free of any prior knowledge of the scene geometry, and no pre-processing step like depth estimation or learning is required, which makes it easy to use and suitable for all types of light field capture.

Furthermore, we proposed to make an experimental study on the compression impact of the full scheme composed of the two proposed methods for light field data acquisition and transmission. The idea is to measure the distortion that the different compression techniques introduce to the original light field. Several datasets were tested, for different compression ratios. The outcome of this study shows that the rate-distortion performance of our full scheme is comparable to that of the disparity-based methods for light field coding. Higher light field quality can be achieved compared to DIBR-based view synthesis methods, but it is still saturated to the quality provided by the reconstruction of the acquired samples.

Future work and perspectives

Several approaches can be considered to extend the works presented in this thesis. The first compression scheme could be improved by making the selection of the angular samples more flexible. Indeed, the selection pattern being inherent to the reconstruction method based the Sparse Fourier transform algorithm (SFFT), it may be of great interest to integrate our frequency selection-based reconstruction method into the base layer of the compression scheme to restore the non-selected views. The reconstruction method should be rethought and adapted to recover angular samples.

Besides, with the imminent delivery of the first standard models of the new codec H266/VVC, a comparison of our compression scheme to this scheme in terms of complexity/ Rate-distortion performance could be considered.

As for the compressive acquisition of light field data, several improvements could be considered. First, a thorough study can be conducted in order to define the optimal sampling pattern that can be applied to each view. A recent work [156] has explored an optimization of the sampling for sparse reconstruction methods by reducing the occurrences of predefined pixel-block structures, to ensure a non-regular and uniform sampling mask. Experiments on 2D images have shown some PSNR improvement. A more theoretical study would still help to find a proved optimal sampling mask for light fields, taking into account both spatial and angular dimensions. As for the compression rate, it would be interesting to conduct a study about the minimum number of samples that ensures an accurate reconstruction of light fields. While, in the compressed sensing theory, a general theoretical threshold [3] has been proposed related to the sparsity and the size of signals, we could extend and adapt this threshold calculation to light field data, by

including their sparsity structure. Moreover, more research could also be conducted towards developing a deep learning-based approach to denoise the 4D Fourier spectrum of a compressively sampled light field.

During this thesis, the research work has focused on light field images, but since more and more light field video contents are being available, our research results could be extended and adapted to light field videos. This could mean adding a temporal dimension to the light field model, and using the inter-frame correlations to ensure a good compressive representation of the light field video.

Publications

F. Hawary, C. Guillemot, D. Thoreau and G. Boisson, "Scalable Light Field Compression Scheme Using Sparse Reconstruction and Restoration", International Conference on Image Processing (ICIP), pp. 3250-3254, 2017.

F. Hawary, G. Boisson, C. Guillemot, and P. Guillotel, “Compressive 4D Light Field Reconstruction Using Orthogonal Frequency Selection", International Conference on Image Processing (ICIP), pp. 3863–3867, 2018.

F. Hawary, G. Boisson, C. Guillemot, and P. Guillotel, "Compressively Sampled Light Field Reconstruction Using Orthogonal Frequency Selection and Refinement", IEEE Transactions on Computational Imaging, (TCI) [submitted].

Patents

F. Hawary, C. Guillemot, and G. Boisson, "Light Field Reconstruction", EP Patent 18306545.7-1209, 2018.

F. Hawary, C. Guillemot, and G. Boisson, "Light Field Reconstruction", EP Patent 18306548.1-1209, 2018.

List of Figures

List of Tables

Bibliography

[1] J. Seiler, M. Jonscher, M. Schoberl, and A. Kaup, "Resampling Images to a Regular Grid From a Non-Regular Subset of Pixel Positions Using Frequency Selective Reconstruction," *IEEE Transactions on Image Processing*, vol. 24, no. 11, pp. 4540–4555, 2015.

[2] S. Mallat and Z. Zhang, "Matching Pursuits with Time-frequency Dictionaries," *IEEE Transactions on Signal Processing*, vol. 41, no. 12, 1993.

[3] E. J. Candès and M. B. Wakin, "An Introduction To Compressive Sampling," *IEEE Signal Processing Magazine*, vol. 25, no. 2, p. 21–30, 2008.

[4] E. H. Adelson and J. R. Bergen, "The plenoptic function and the elements of early vision," *Computational models of visual processing*, vol. 1, no. 2, 1991.

[5] S. J. Gortler, R. Grzeszczuk, R. Szeliski, and M. F. Cohen, "The lumigraph," *Proc. of the 23rd Annual Conference on Computer Graphics and Interactive Techniques*, p. 43–54, 1996.

[6] B. H. Bolles, R. and D. Marimont, "Epipolar-plane image analysis: An approach to determining structure from motion," *International Journal of Computer Vision*, vol. 1, p. 7–55, 1987.

[7] S. A. Benton, "Survey of holographic stereograms," *In 26th Annual Technical Symposium*, p. 15–19, 1983.

[8] B. Wilburn, N. Joshi, V. Vaish, M. Levoy, and M. Horowitz, "High-speed videography using a dense camera array," *CVPR*, vol. 2, pp. 294–301, 2004.

[9] J. C. Yang, "A light field camera for image based rendering," *PhD thesis, Massachusetts Institute of Technology*, 2000.

[10] N. Sabater, G. Boisson, B. Vandame, P. Kerbiriou, F. Babon, M. Hog, R. Gendrot, T. Langlois, O. Bureller, A. Schubert, and V. Allié, "Dataset and Pipeline for Multiview Light-Field Video," *IEEE Conference on Computer Vision and Pattern Recognition Workshops (CVPRW)*, p. 1743–1753, 2017.

[11] L. Dabala, M. Ziegler, P. Didyk, F. Zilly, J. Keinert, K. Myszkowski, H.-P. Seidel, P. Rokita, and T. Ritschel, "Efficient Multi-image Correspondences for Online Light Field Video Processing," *Computer Graphics Forum*, vol. 35, no. 7, pp. 401–410, 2016.

[12] B. Wilburn, N. Joshi, V. Vaish, E. Talvala, E. Antunez, A. Barth, A. Adams, M. Horowitz, and M. Levoy, "High performance imaging using large camera arrays," *ACM Trans. on Graphics (TOG)*, vol. 24, no. 3, p. 765, 2005.

[13] E. H. Adelson and J. Y. A. Wang, "Single lens stereo with a plenoptic camera," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, no. 2, pp. 99–106, 1992.

[14] R. Ng, M. Levoy, M. Brédif, G. Duval, M. Horowitz, and P. Hanrahan, "Light field photography with a hand-held plenoptic camera," *Computer Science Technical Report*, vol. 2, no. 11, pp. 1–11, 2005.

[15] F. Durand, N. Holzschuch, C. Soler, E. Chan, and F. X. Sillion, "A Frequency Analysis of Light Transport," *ACM Transactions on Graphics*, vol. 24, no. 3, p. 1115–1126, 2005.

[16] Z. Zhang and M. Levoy, "Wigner Distributions and How They Relate to The Light Field," *Proc. IEEE Int. Conf. Comput. Photography*, p. 1–10, 2009.

[17] R. Ng, "Fourier slice photography," *ACM Transactions on Graphics*, vol. 24, no. 3, p. 735, 2005.

[18] A. Levin, S. W. Hasinoff, P. Green, F. Durand, and W. T. Freeman, "4D Frequency Analysis of Computational Cameras for Depth of Field Extension," *ACM Transactions on Graphics*, vol. 28, no. 3, pp. 97:1–97:14, 2009.

[19] A. Veeraraghavan, R. Raskar, A. Agrawal, A. Mohan, and J. Tumblin, "Dappled photography: mask enhanced cameras for heterodyned light fields and coded aperture refocusing," *ACM Transactions on Graphics*, 2007.

[20] A. Levin and F. Durand, "Linear view synthesis using a dimensionality gap light field prior," *Proceedings of the IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, pp. 1831–1838, 2010.

[21] K. Marwah, G. Wetzstein, Y. Bando, and R. Raskar, "Compressive Light Field Photography using Overcomplete Dictionaries and Optimized Projections," *ACM Trans. Graph. (Proc. SIGGRAPH)*, vol. 32, no. 4, pp. 1–11, 2013.

[22] L. Shi, H. Hassanieh, A. Davis, D. Katabi, and F. Durand, "Light field reconstruction using sparsity in the continuous Fourier domain," *ACM Transactions on Graphics*, vol. 34, no. 1, pp. 1–13, 2014.

[23] S. Wanner and B. Goldluecke, "Variational light field analysis for disparity estimation and super-resolution," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 36, no. 3, pp. 606–619, 2014.

[24] J. Li, M. Lu, and Z.-N. Li, "Continuous depth map reconstruction from light fields," *IEEE Transactions on Image Processing*, vol. 24, no. 11, p. 3257–3265, 2015.

[25] C. Kim, H. Zimmer, Y. Pritch, A. Sorkine-Hornung, and M. Gross, "Scene reconstruction from high spatio-angular resolution light fields," *ACM Transactions on Graphics*, vol. 32, no. 4, 2013.

[26] H.-G. Jeon, J. Park, G. Choe, J. Park, Y. Bok, Y.-W. Tai, and I. S. Kweon, "Accurate depth map estimation from a lenslet light field camera," *Proceedings on Conference Computer Vision and Pattern Recognition*, pp. 1547–1555, 2015.

[27] T.-C. Wang, A. A. Efros, and R. Ramamoorthi, "Occlusion-aware depth estimation using light-field cameras," *IEEE Proceedings on International Conference Computer Vision*, p. 3487–3495, 2015.

[28] H. Zhu, Q. Wang, and J. Yu, "Occlusion-model guided anti-occlusion depth estimation in light field," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, p. 965–978, 2017.

[29] C. Chen, H. Lin, Z. Yu, S. B. Kang, and J. Yu, "Light field stereo matching using bilateral statistics of surface cameras," *in Proc. IEEE Conference Computer Vision Pattern Recognition*, p. 1518–1525, 2014.

[30] M. W. Tao, S. Hadap, J. Malik, and R. Ramamoorthi, "Depth from combining defocus and correspondence using light-field cameras," *Proc. IEEE Conference Computer Vision and Pattern Recognition*, p. 673–680, 2013.

[31] M. Tao, P. P. Srinivasan, J.Malik, S. Rusinkiewicz, and R. Ramamoorthi, "Depth from shading, defocus, and correspondence using light-field angular coherence," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, p. 1940–1948, 2015.

[32] W. Williem and I. K. Park, "Robust light field depth estimation for noisy scene with occlusion," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, p. 4396–4404, 2016.

[33] S. Heber and T. Pock, "Convolutional networks for shape from light field," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, p. 3746–3754, 2016.

[34] O. Jonhannsen, A. Sulc, and B. Goldluecke, "What sparse light field coding reveals about scene structure," *Proc. IEEE Conference on Computer Vision and Pattern Recognition*, p. 3262–3270, 2016.

[35] Z. Xiong, L. Wang, H. Li, D. Liu, and F. Wu, "Snapshot hyperspectral light field imaging," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, p. 3270–3278, 2017.

[36] H. Rueda, C. Fu, D. L. Lau, and G. R. Arce, "Spectral-TOF compressive snapshot camera: Towards hyperspectral+depth imagery," *IEEE J. Sel. Topics Signal Process.*, vol. 11, no. 7, p. 992–1003, 2017.

[37] G. Todor and L. Andrew, "Super-Resolution with Plenoptic Camera 2.0," *Adobe Systems*, 2009.

[38] T. E. Bishop and P. Favaro, "The light field camera: Extended depth of field, aliasing, and super-resolution," *IEEE Trans. on Pattern Analysis & Machine Intelligence*, vol. 34, no. 5, p. 972–986, 2012.

[39] K. Mitra and V. Ashok, "Light field denoising, light field superresolution and stereo camera based refocusing using a GMM light field patch prior," *Proc. IEEE Comput. Soc. Conf.Comput. Vis. Pattern Recognit. Workshops*, p. 22–28, 2012.

[40] V. Boominathan, K. Mitra, and A. Veeraraghavan, "Improving Resolution and Depth-of-field of Light Field Cameras Using a Hybrid Imaging System," *Proc. IEEE Int. Conf. Comput. Photography*, p. 1–10, 2014.

[41] Y. Wang, Y. Liu, W. Heidrich, and Q. Dai, "The Light Field Attachment: Turning a DSLR into a Light Field Camera Using a Low Budget Camera Ring," *IEEE Trans. Vis. Comput. Graph.*, vol. 23, no. 10, pp. 2357–2364, 2016.

[42] Y. Yoon, H. G. Jeon, D. Yoo, J. Y. Lee, and I. S. Kweon, "Learning a Deep Convolutional Network for Light-Field Image Super-Resolution," *Proceedings of the IEEE International Conference on Computer Vision*, pp. 57–65, 2015.

[43] R. A. Farrugia, C. Galea, and C. Guillemot, "Super-resolution of light field images using linear subspace projection of patch-volumes," *IEEE Journal on Selected Topics Signal Processing*, vol. 11, no. 7, p. 1058–1071, 2017.

[44] T. Georgiev, K. C. Zheng, B. Curless, D. Salesin, S. Nayar, and C. Intwala, "Spatio-angular resolution tradeoffs in integral photography," *Proc. Eurograph. Symp. Rendering*, p. 263–272, 2006.

[45] G. Chaurasia, S. Duchêne, O. Sorkine-Hornung, , and G. Drettakis, "Depth synthesis and local warps for plausible image-based navigation," *ACM Transactions on Graphics*, vol. 32, 2013.

[46] F. D. S. Pujades and B. Goldluecke, "Bayesian view synthesis and image-based rendering principles," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, p. 3906–3913, 2014.

[47] J. P. J. Flynn, I. Neulander and N. Snavely, "DeepStereo: Learning to predict new views from the world's imagery," *Proc. IEEE Conf. Comput. Vis. Pattern Recognit.*, p. 5515–5524, 2015.

[48] N. K. Kalantari, T. C. Wang, and R. Ramamoorthi, "Learning-Based View Synthesis for Light Field Cameras," *ACM Transactions on Graphics (Proc. of SIGGRAPH Asia)*, vol. 35, no. 6, 2016.

[49] P. P. Srinivasan, T. Wang, A. Sreelal, R. Ramamoorthi, and R. Ng, "Learning to Synthesize a 4D RGBD Light Field from a Single Image," *Proc. of the IEEE International Conference on Computer Vision (ICCV)*, pp. 2262–2270, 2017.

[50] S. Vagharshakyan, R. Bregovic, and A. Gotchev, "Light field reconstruction using Shearlet transform," *IEEE Transactions on Pattern Analysis and Machine Intelligence*, vol. 40, no. 1, pp. 133–147, 2018.

[51] ——, "Accelerated Shearlet-Domain Light Field Reconstruction," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1082–1091, 2017.

[52] M. Levoy and P. Hanrahan, "Light field rendering," *Proceedings of SIGGRAPH*, pp. 31–42, 1996.

[53] M. Leonard and B. Gary, "Plenoptic modeling: An image-based rendering system," *Proc. of the 22nd Annual Conference on Computer Graphics and Interactive Techniques*, p. 39–46, 1995.

[54] J.-X. Chai, X. Tong, S.-C. Chan, and H.-Y. Shum, "Plenoptic sampling," *Proc. of the 27nd Annual Conference on Computer Graphics and Interactive Techniques*, p. 307–318, 2000.

[55] Z. Lin and H.-Y. Shum, "A geometric analysis of light field rendering," *International Journal of Computer Vision*, vol. 58, no. 2, p. 121–138, 2004.

[56] C. Buehler, M. Bosse, L. McMillan, S. Gortler, and M. Cohen, ""Unstructured lumigraph rendering," *Proc. of the 28th Annual Conference on Computer Graphics and Interactive Techniques*, pp. 425–432, 2001.

[57] L. Yatziv, G. Sapiro, and M. Levoy, "Light field completion," *Proc. IEEE Int. Conf. Image Process.*, p. 1787–1790, 2004.

[58] M. Levoy, B. Chen, V. Vaish, M. Horowitz, I. McDowall, and M. Bolas, "Synthetic Aperture Confocal Imaging," *ACM Trans. Graph.*, vol. 23, p. 825–834, 2004.

[59] M. Hog, N. Sabater, and C. Guillemot, "Light field segmentation using a ray-based graph structure," *Proc. IEEE Eur. Conf. Comput. Vis*, pp. 35–50, 2016.

[60] K. Yucer, A. Sorkine-Hornung, O.Wang, and O. Sorkine-Hornung, "Efficient 3D object segmentation from densely sampled light fields with applications to 3D reconstruction," *ACM Transactions on Graphics*, vol. 35, no. 5, pp. 249–257, 2016.

[61] Q. Z. H. Zhu and Q. Wang, "4D light field super-pixel and segmentation," *Proc. IEEE Conf. Computer Vision and Pattern Recognition*, p. 6384–6392, 2017.

[62] A. S. O. Johannsen and B. Goldluecke, "Variational separation of light field layers," *Proc. Vis., Model. Vis.*, p. 135–142, 2015.

[63] Y. Xu, H. Nagahara, A. Shimada, and R. Taniguchi, "Transcut: Transparent object segmentation from a light field image," *Proc. IEEE Internationl Conference on Computer Vision*, p. 3442–3450, 2015.

[64] Z. Pei, Y. Zhang, T. Yang, X. Zhang, and Y.-H. Yang, "A novel multi-object detection method in complex scene using synthetic aperture imaging," *Pattern Recognition*, vol. 45, p. 1637–1658, 2012.

[65] R. Gross, I. Matthews, and S. Baker, "Appearance-based face recognition and light fields," *IEEE Trans. Pattern Aanl. Mach. Intell.*, vol. 26, no. 4, p. 449–465, 2004.

[66] R. Raghavendra, B. Yang, K. B. Raja, and C. Busch, "A new perspective—face recognition with light field camera," *Proc. Int. Conf. Biometrics*, pp. 1–8, 2013.

[67] T.-C. Wang, J.-Y. Zhu, E. Hiroaki, M. Chandraker, A. A. Efros, and R. Ramamoorthi, "A 4D light-field dataset and CNN architectures for material recognition," *Proc. IEEE Eur. Conf. Computer Vision*, p. 121–138, 2016.

[68] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Scalable coding of plenoptic images by using a sparse set and disparities," *IEEE Trans. on Image Processing*, vol. 25, no. 1, pp. 80–91, 2016.

[69] M. Rizkallah, T. Maugey, C. Yaacoub, and C. Guillemot, "Impact of light field compression on focus stack and extended focus images," *EUSIPCO*, pp. 898–902, 2016.

[70] Q. Fu, Z. Zhou, Y. Yuan, and B. Xiangli, "Image quality evaluation of light field photography," *Image Quality System Performance*, vol. 7867, no. 4, p. 357–366, 2011.

[71] Z.Wang, A. C. Bovik, H. R. Sheikh, and E. P. Simoncelli, "Image quality assessment: From error visibility to structural similarity," *IEEE Trans. Image Process.*, vol. 13, no. 4, p. 600–612, 2004.

[72] H. Shidanshidi, F. Safaei, and W. Li, "A quantitative approach for comparison and evaluation of light field rendering techniques," *IEEE Int. Conf. Multimedia Expo*, vol. 50, no. 4, pp. 1–4, 2011.

[73] V. K. Adhikarla, M. Vinkler, D. Sumin, R. K. Mantiuk, K. Myszkowski, H.-P. Seidel, and P. Didyk, "Towards a Quality Metric for Dense Light Fields," *IEEE Conference on Computer Vision and Pattern Recognition*, p. 58–67, 2017.

[74] A. Aggoun, "A 3D DCT compression algorithm for omnidirectional integral images," *in Proc. 2006 IEEE ICASSP*, vol. 2, p. 517–520, 2006.

[75] M. B. de Carvalho, M. P. Pereira, C. L. Pagliari, F. Pereira, G. Alves, V. Testoni, and E. A. B. Silva, "a 4D DCT-Based Lenslet Light Field Codec," *ICIP*, pp. 435–439, 2018.

[76] P. Lalonde and A. Fournier, "Interactive rendering of wavelet projected light fields," in *Proc. of the Conference on Graphics Interface*, 1999, pp. 107–114.

[77] I. Peter and W. Straßer, "The wavelet stream - progressive transmission of compressed light field data," in *IEEE Visualization 1999 Late Breaking Hot Topics.* IEEE Computer Society, 1999, pp. 69–72.

[78] A. Aggoun, "Compression of 3D integral images using 3D wavelet transform," *J. Display Technol.*, vol. 7, no. 11, p. 586–592, 2011.

[79] M. Magnor, A. Endmann, and B. Girod, "Progressive compression and rendering of light fields," *Vision, Modeling and Visualization*, pp. 199–203, 2000.

[80] C. L. Chang, X. Zhu, P. Ramanathan, and B. Girod, "Light field compression using disparity-compensated Lifting & Shape Adaptation," *IEEE Transactions on Image Processing*, vol. 1, no. 4, pp. I373–I376, 2006.

[81] X. Dong, D. Qionghai, and X. Wenli, "Light field compression based on prediction propagating and wavelet packet," *ICIP*, vol. 5, pp. 3515–3518 Vol. 5, 2004.

[82] T. Sakamoto, K. Kodama, and T. Hamamoto, "A study on efficient compression of multi-focus images for dense light-field reconstruction," *VCIP*, 2012.

[83] E. Elharar, A. Stern, O. Hadar, and B. Javidi, "A hybrid compression method for integral images using discrete wavelet transform and discrete cosine transform," *J. Display Technol.*, vol. 3, no. 3, p. 321–325, 2007.

[84] D. Lelescu and F. Bossen, "Representation and coding of light field data," *Graph. Models*, vol. 66, no. 4, pp. 203–225, Jul. 2004.

[85] G. J. Sullivan, J. R. Ohm, W. J. Han, and T. Wiegand, "Overview of the high efficiency video coding (HEVC) standard," *IEEE Trans. on Circuits and Systems for Video Technology*, vol. 22, no. 12, pp. 1649–1668, 2012.

[86] C. Conti, L. D. Soares, and P. Nunes, "HEVC-based 3D holoscopic video coding using self-similarity compensated prediction," *Signal Process.: Image Commun.*, vol. 42, p. 59–78, 2016.

[87] C. Conti, P. Nunes, and L. D. Soares, "Hevc-based light field image coding with bi-predicted self-similarity compensation," *Proc. IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, p. 1–4, 2016.

[88] R. Monteiro, L. Lucas, C. Conti, P. Nunes, N. Rodrigues, S. Faria, C. Pagliari, E. da Silva, and L. Soares, "Light field HEVC-based image coding using locally linear embedding and self-similarity compensated prediction," *2016 IEEE International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–4, 2016.

[89] D. Liu, P. An, R. Ma, C. Yang, and L. Shen, "3D Holoscopic Image Coding Scheme Using HEVC with Gaussian Process Regression," *Signal Process. Image Commun.*, vol. 47, p. 438–451, 2016.

[90] Y. Li, M. Sjöström, R. Olsson, and U. Jennehag, "Coding of Focused Plenoptic Contents by Displacement Intra Prediction," *IEEE Trans. Circuits Syst. Video Techn.*, vol. 26, no. 7, pp. 1308–1319, 2016.

[91] A. Vetro, T. Wiegand, and G. J. Sullivan, "Overview of the Stereo and Multiview Video Coding Extensions of the H.264/MPEG-4 AVC Standard," *Proceedings of the IEEE*, vol. 99, no. 4, p. 626–642, 2011.

[92] S. Shi, P. Gioia, and G. Madec, "Efficient compression method for integral images using multi-view video coding," *18th IEEE Int. Conf. Image Processing(ICP)*, pp. 137—140, 2011.

[93] J. Dick, H. Almeida, L. Soares, and P. Nunes, "3D Holoscopic Video Coding Using MVC," *IEEE EUROCON - International Conference on Computer as a Tool*, p. 1–4, 2011.

[94] G. Tech, Y. Chen, K. Müller, J. R. Ohm, A. Vetro, and Y. K. Wang, "Overview of the multiview and 3D extensions of high efficiency video coding," *IEEE Transactions on Circuits and Systems for Video Technology*, vol. 26, no. 1, pp. 35–49, 2016.

[95] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet-Popescu, and F. Dufaux, "Full Parallax 3D Video Content Compression," *Novel 3D Media Technologies*, 2015.

[96] G. Wang, W. Xiang, M. Pickering, and C. W. Chen, "Light field multi-view video coding with two-directional parallel inter-view prediction," *IEEE Trans. on Image Processing*, vol. PP, no. 99, pp. 5104–5117, 2016.

[97] T. Wiegand, G. J. Sullivan, G. Bjøntegaard, and A. Luthra, "Overview of the H.264/AVC video coding standard," *IEEE Trans. Circuits and Systems for Video Technology*, vol. 13, no. 7, pp. 560–576, 2003.

[98] R. Olsson, "Empirical rate-distortion analysis of JPEG 2000 3D and H.264/AVC coded integral imaging based 3D-images," *3DTV-Conference Proceedings*, pp. 113–116, 2008.

[99] A. Vieira, H. Duarte, C. Perra, L. Tavora, and P. Assuncao, "Data formats for high efficiency coding of Lytro-Illum light fields," *5th International Conference on Image Processing, Theory, Tools and Applications*, pp. 494–497, 2015.

[100] I. Viola, M. Řeřábek, T. Bruylants, P. Schelkens, F. Pereira, and T. Ebrahimi, "Objective and subjective evaluation of light field image compression algorithms," *Picture Coding Symposium (PCS)*, 2016.

[101] R. Olsson, M. Sjöström, and Y. Xu, "A Combined Pre-processing and H.264-compression Scheme for 3D Integral Images," *ICIP*, pp. 513–516, 2006.

[102] F. Dai, J. Zhang, Y. Ma, and Y. Zhang, "Lenselet image compression scheme based on subaperture images streaming," *Proc. in International Conference on Image Processing (ICIP)*, vol. 2015-Decem, pp. 4733–4737, 2015.

[103] S. Zhao, Z. Chen, K. Yang, and H. Huang, "Light field image coding with hybrid scan order," *30th Anniversary of Visual Communication and Image Processing (VCIP)*, pp. 1–4, 2016.

[104] X. Jiang, M. Le Pendu, R. A. Farrugia, S. Hemami, and C. Guillemot, "Homography-based low rank approximation of light fields for compression," *ICASSP*, 2017.

[105] L. Li, Z. Li, B. Li, D. Liu, and H. Li, "Pseudo-Sequence-Based 2-D Hierarchical Coding Structure for Light-Field Image Compression," *IEEE Journal on Selected Topics in Signal Processing*, vol. 11, no. 7, pp. 1107–1119, 2017.

[106] C. Perra and P. Assuncao, "High efficiency coding of light field images based on tiling and pseudo-temporal data arrangement," *2016 IEEE International Conference on Multimedia and Expo Workshop (ICMEW)*, 2016.

[107] X. Jiang, M. Le Pendu, R. A. Farrugia, and C. Guillemot, "Light Field Compression with Homography-based Low Rank Approximation," *IEEE Journal of Selected Topics in Signal Processing*, vol. 11, pp. 1132–1145, 2017.

[108] X. Zhang, H. Wang, and T. Tian, "Light Field Image Coding with Disparity Correlation Based Prediction," *IEEE Fourth International Conference on Multimedia Big Data (BigMM)*, pp. 1–6, 2018.

[109] W. Ahmad, R. Olsson, and M. Sjöström, "Towards a Generic Compression Solution for Densely and Sparsely Sampled Light Field Data," *ICIP*, pp. 654–658, 2018.

[110] A. Dricot, J. Jung, M. Cagnazzo, B. Pesquet, and F. Dufaux, "Integral images compression scheme based on view extraction," *23rd European Signal Processing Conference, EUSIPCO*, pp. 101–105, 2015.

[111] C. Conti, P. Nunes, and L. D. Soares, "Inter-layer prediction scheme for scalable 3-D holoscopic video coding," *IEEE Signal Processing Letters*, vol. 20, no. 8, pp. 819–822, 2013.

[112] C.-K. Liang, T.-H. Lin, B.-Y. Wong, C. Liu, and H. H. Chen, "Programmable aperture photography: multiplexed light field acquisition," in *Proc. of ACM SIGGRAPH*, vol. 27, 2008, pp. 1–10.

[113] Z. Xu and E. Y. Lam, "A high-resolution light field camera with dual-mask design," *In Proc. SPIE*, vol. 8500, p. 85000U, 2012.

[114] Q. D. Zhoutong Zhang, Yebin Lin, "Light Field from Micro-baseline Image Pair," *CVPR*, 2015.

[115] Y. Yagi, K. Takahashi, T. Fujii, T. Sonoda, and H. Nagahara, "PCA-coded aperture for light field photography," *IEEE ICIP*, no. 2, pp. 3031–3035, 2017.

[116] ——, "Designing Coded Aperture Camera Based on PCA and NMF for Light Field Acquisition," *IEICE Transactions on Information and Systems*, vol. 101, pp. 2190–2200, 2018.

[117] D. L. Donoho, "Compressed sensing," *IEEE Transactions on Information Theory*, vol. 52, no. 4, p. 1289–1306, 2006.

[118] Y. C. Eldar and G. Kutyniok, "Compressed Sensing: Theory and Applications," *Cambridge University Press*, 2012.

[119] E. J. Candès, N. Braun, and M. B. Wakin, "Sparse Signal and Image Recovery from Compressive Samples," in *Proc. of the Int. Symposium on Biomedical Imaging: From Nano to Macro*, 2007, pp. 976–979.

[120] J. R. E. J. Candès and T. Tao, "Robust Uncertainty Principles: Exact Signal Reconstruction From Highly Incomplete Frequency Information," *IEEE Transactions on Information Theory*, vol. 52, no. 2, p. 489–509, 2006.

[121] E. J. Candès and T. Tao, "Near-optimal signal recovery from random projections: Universal encoding strategies?" *IEEE Transactions on Information Theory*, vol. 52, no. 12, p. 5406–5425, 2006.

[122] A. Ashok and M. A. Neifeld, "Compressive Light Field Imaging," *In Proc. SPIE*, vol. 7690, p. 76900Q, 2010.

[123] S. D. Babacan, R. Ansorge, M. Luessi, P. R. Mataran, R. Molina, and A. K. Katsaggelos, "Compressive Light Field Sensing," *IEEE Transactions on Image Processing*, vol. 21, no. 12, pp. 4746–4757, 2012.

[124] Y. P. Wang, L. C. Wang, D. H. Kong, and B. C. Yin, "High-Resolution Light Field Capture with Coded Aperture," *IEEE Transactions on Image Processing*, vol. 24, no. 12, pp. 5609–5618, 2015.

[125] E. Miandji, J. Kronander, and J. Unger, "Compressive Image Reconstruction in Reduced Union of Subspaces," *Computer Graphics Forum*, vol. 34, no. 2, pp. 33–44, 2015.

[126] C. G. Ehsan Miandji, Jonas Unger, "Multi-shot single sensor light field camera using a color coded mask," *EUSIPCO*, pp. 1–5, 2018.

[127] M. Gupta, A. Jauhari, K. Kulkarni, S. Jayasuriya, A. Molnar, and P. Turaga, "Compressive Light Field Reconstructions Using Deep Learning," *CVPRW*, pp. 1277–1286, 2017.

[128] A. K. Vadathya, S. Cholleti, G. Ramajayam, V. Kanchana, and K. Mitra, "Learning Light Field Reconstruction from a Single Coded Image," *ACPR*, 2017.

[129] O. Nabati, D. Mendlovic, and R. Giryes, "Fast and Accurate Reconstruction of Compressed Color Light Field," *IEEE International Conference on Computational Photography (ICCP)*, pp. 1–11, 2018.

[130] H. Hassanieh, P. Indyk, D. Katabi, and E. Price, "Simple and practical algorithm for sparse fft," *In Proc. of the 23rd Annual ACM-SIAM Symposium on Discrete Algorithms*, p. 1183–1194, 2012.

[131] B. Ghazi, H. Hassanieh, P. Indyk, D. Katabi, E. Price, and L. Shi, "Sample-optimal average-case sparse Fourier Transform in two dimensions," *2013 51st Annual Allerton Conference on Communication, Control, and Computing, Allerton 2013*, pp. 1258–1265, 2013.

[132] K. Meisinger and A. Kaup, "Spatial Error Concealment of Corrupted Image Data Using Frequency Selective Extrapolation," *International Conference on Acoustics, Speech, and Signal Processing (ICASSP)*, pp. III–209 – III–212, 2004.

[133] ——, "Minimizing a Weighted Error Criterion for Spatial Error Concealment of Missing Image Data," *IEEE International Conference on Image Processing (ICIP)*, vol. 2, no. 3, pp. 813–816, 2004.

[134] J. Seiler, K. Meisinger, and A. Kaup, "Orthogonality Deficiency Compensation for Improved Frequency Selective Image Extrapolation," *Proc. in Picture Coding Symposium*, no. 1, pp. 1–4, 2007.

[135] J. Seiler, H. Lakshman, and A. Kaup, "Spatio-temporal Prediction in Video Coding by Best Approximation," *PCS*, 2009.

[136] A. Kaup and T. Aach, "Coding of segmented images using shape-independent basis functions," *IEEE Transactions on Image Processing*, vol. 7, no. 7, pp. 937–947, 1998.

[137] J. Seiler and A. Kaup, "Motion Compensated Frequency Selective Extrapolation for Error Concealment in Video Coding," *EUSIPCO*, 2008.

[138] ——, "Content-adaptive Motion Compensated Frequency Selective Extrapolation for Error Concealment in Video Communication," *ICIP*, no. 1, pp. 445–448, 2010.

[139] ——, "Optimizd and Parallelized Processing Order for Improved FSE," *EUSIPCO*, pp. 269–273, 2011.

[140] J. Seiler, S. Scholl, W. Schnurrer, and A. Kaup, "Optimized Processing Order for 3D Hole Filling in Video Sequences Using Frequency Selective Extrapolation," *Picture Coding Symposium (PCS)*, 2016.

[141] M. Jonscher, J. Seiler, and A. Kaup, "Texture-Dependent Frequency Selective Reconstruction of Non-Regularly Sampled Images," *Picture Coding Symposium (PCS)*, no. 1, pp. 1–5, 2016.

[142] K. Meisinger and A. Kaup, "2D Frequency Selective Extrapolation for Spatial Error Concealment in H.264/AVC Video Coding," *IEEE International Conference on Image Processing (ICIP)*, pp. 2233–2236, 2006.

[143] U. Fecker, J. Seiler, and A. Kaup, "4-D Frequency Selective Extrapolation for Error Concealment in Multi-view Video," *Proc. of the 10th Workshop on Multimedia Signal Processing (MMSP)*, pp. 267–272, 2008.

[144] J. Koloda, J. Seiler, and A. Kaup, "Frequency-Selective Mesh-to-Grid Resampling for Image Communication," *IEEE Transactions on Multimedia*, vol. 19, no. 8, pp. 1689–1701, 2017.

[145] G. J. Sullivan, J. M. Boyce, Y. Chen, J. Ohm, C. A. Segall, and A. Vetro, "Standardized Extensions of High Efficiency Video Coding (HEVC)," *IEEE Journal on Selected Topics in Signal Processing*, vol. 7, no. 6, pp. 1001–1016, 2013.

[146] H. Hassanieh, "The Sparse Fourier Transfor: theory and pratice," *PhD thesis*, 2016.

[147] C. Barnes, E. Shechtman, A. Finkelstein, and D. B. Goldman, "PatchMatch: A Randomized Correspondence Algorithm for Structural Image Editing," *ACM Trans. on Graphics*, vol. 28, no. 3, p. 1, 2009.

[148] D. Simakov, Y. Caspi, E. Shechtman, and M. Irani, "Summarizing visual data using bidirectional similarity," in *IEEE (CVPR)*, 2008.

[149] M. Rerabek and T. Ebrahimi, "New Light Field Image Dataset," *8th International Workshop on Quality of Multimedia Experience (QoMEX)*, 2016.

[150] G. Bjontegaard, "Calculation of average PSNR differences between RD-curves," *Doc. VCEG-M33, ITU-T VCEG Meeting*, 2001.

[151] R. Ng, "Digital light field photography," *Stanford University*, pp. 1–203, 2006.

[152] P. Marziliano, F. Dufaux, S. Winkler, and T. Ebrahimi, "A no-reference perceptual blur metric," *Proc. of the International Conference on Image Processing (ICIP)*, vol. 3, pp. 8–11, 2002.

[153] R. S. Overbeck, D. Erickson, D. Evangelakos, M. Pharr, and P. Debevec, "A System for Acquiring, Processing, and Rendering Panoramic Light Field Stills for Virtual Reality," *ACM Transactions on Graphics*, vol. 37, no. 6, 2018.

[154] X. Jiang, M. Le Pendu, and C. Guillemot, "Light field compression using depth image based view synthesis," *IEEE International Conference on Multimedia and Expo Workshops (ICMEW)*, vol. 694122, pp. 19–24, 2017.

[155] P. Weinzaepfel, J. Revaud, Z. Harchaoui, and C. Schmid, "DeepFlow: Large displacement optical flow with deep matching," *IEEE InteRnational Conference on Computer Vision (ICCV)*, pp. 1385–1392, 2013.

[156] S. Grosche, J. Seiler, and A. Kaup, "Iterative Optimization of Quarter Sampling Masks for Non-Regular Sampling Sensors," *IEEE International Conference on Image Processing*, pp. 26–30, 2018.

Titre : Compression et Acquisition Comprimée de Champs de Lumière

Mots clés : Compression, Imagerie de champs de lumière, acquisition comprimée, traitement d'image

Résumé : En capturant une scène à partir de plusieurs points de vue, un champ de lumière fournit une représentation riche de la géométrie de la scène, ce qui permet une variété de nouvelles applications de post-capture ainsi que des expériences immersives. L'objectif de cette thèse est d'étudier la compressibilité des contenus de type champ de lumière afin de proposer de nouvelles solutions pour une imagerie de champs lumière à plus haute résolution. Deux aspects principaux ont été étudiés à travers ce travail.
Les performances en compression sur les champs lumière des schémas de codage actuels étant encore limitées, il est nécessaire d'introduire des approches plus adaptées aux structures des champs de lumière. Nous proposons un schéma de compression comportant deux couches de codage. Une première couche encode uniquement un sous-ensemble de vues d'un champ de lumière et reconstruit les vues restantes via une méthode basée sur la parcimonie. Un codage résiduel améliore ensuite la qualité finale du champ de lumière décodé.
Avec les moyens actuels de capture et de stockage, l'acquisition d'un champ de lumière à très haute résolution spatiale et angulaire reste impossible, une alternative consiste à reconstruire le champ de lumière avec une large résolution à partir d'un sous-ensemble d'échantillons acquis. Nous proposons une méthode de reconstruction automatique pour restaurer un champ de lumière échantillonné. L'approche utilise la parcimonie du champs de lumière dans le domaine de Fourier. Aucune estimation de la géométrie de la scène n'est nécessaire, et une reconstruction précise est obtenue même avec un échantillonnage assez réduit. Une étude supplémentaire du schéma complet, comprenant les deux approches proposées est menée afin de mesurer la distorsion introduite par les différents traitements. Les résultats montrent des performances comparables aux méthodes de synthèse de vues basées sur la l'estimation de profondeur.

Title: Light Field Image Compression and Compressive Acquisition

Keywords: Compression, Light field imaging, Compressive sensing, Image Processing

Abstract: By capturing a scene from several points of view, a light field provides a rich representation of the scene geometry that brings a variety of novel post-capture applications and enables immersive experiences. The objective of this thesis is to study the compressibility of light field contents in order to propose novel solutions for higher-resolution light field imaging. Two main aspects were studied through this work. The compression performance on light fields of the actual coding schemes still being limited, there is need to introduce more adapted approaches to better describe the light field structures. We propose a scalable coding scheme that encodes only a subset of light field views and reconstruct the remaining views via a sparsity-based method. A residual coding provides an enhancement to the final quality of the decoded light field.
Acquiring very large-scale light fields is still not feasible with the actual capture and storage facilities, a possible alternative is to reconstruct the densely sampled light field from a subset of acquired samples. We propose an automatic reconstruction method to recover a compressively sampled light field, that exploits its sparsity in the Fourier domain. No geometry estimation is needed, and an accurate reconstruction is achieved even with very low number of captured samples.
A further study is conducted for the full scheme including a compressive sensing of a light field and its transmission via the proposed coding approach. The distortion introduced by the different processing is measured. The results show comparable performances to depth-based view synthesis methods.